

Ana Rita Mendanha Simão

Uma caracterização combinatória das árvores filogenéticas



**Departamento de Matemática
Faculdade de Ciências da Universidade do Porto
2014**

Ana Rita Mendanha Simão

Uma caracterização combinatória das árvores filogenéticas



*Tese submetida à Faculdade de Ciências da
Universidade do Porto para obtenção do grau de Mestre
em Matemática*

Orientadores: Prof.^a Doutora Maria Leonor Nogueira Coelho Moreira
Prof. Doutor Samuel António de Sousa Dias Lopes

Departamento de Matemática
Faculdade de Ciências da Universidade do Porto
2014

Agradecimentos

Concluindo este meu trabalho, é justo prestar homenagem e os meus sinceros agradecimentos a todos os que pela sua ação, direta ou indiretamente, o tornaram viável. Em primeiro lugar, aos meus orientadores, Professora Doutora Maria Leonor Moreira e Professor Doutor Samuel Dias Lopes, pelos seus ensinamentos, pela pertinência nas sugestões e críticas e pela enorme disponibilidade demonstrada. De facto, a sensibilidade evidenciada ao longo deste percurso, não só engrandeceu o meu trabalho como me ajudou a perceber uma vertente humana capaz de me apontar um trilho de crescimento afetivo. Na verdade, da simbiose entre o trabalho propriamente dito e da componente emocional colocada ao dispor desse mesmo trabalho, resulta sempre uma enorme gratificação pessoal e profissional.

À minha família que sempre me apoiou em todos os meus projetos, proporcionando sempre condições ideais para que os pudesse realizar. Inexcedível nesse apoio, colocou sempre como prioridade a concretização deste meu projeto; convicta de que o tempo que lhe dediquei serviria para enfatizar uma componente que me enriqueceria, sensibilizada, pude observar que esse apoio ultrapassou os parâmetros da razoabilidade tendo, assim, contribuído para um solidificar de laços que, gradativamente, foram estreitados mercê da confiança e do incentivo de que fui alvo.

Às minhas colegas de mestrado, Carla Azevedo e Elisa Silveira, que me apoiaram ao longo destes dois anos e que foram uma preciosa ajuda para a concretização deste mestrado. De facto, em espírito de colaboração e de ajuda interpessoal, pude enriquecer o meu trabalho colocando a tónica na análise crítica que, de forma positiva, me ajudou a desbravar caminhos que, eventualmente, pudessem apresentar-se mais obscuros.

À Vânia Pinheiro pela paciência com que leu este trabalho e pela disponibilidade sempre demonstrada. Atendendo ao facto de que os seus múltiplos afazeres poderiam coartar esse mesmo apoio, esta minha caminhada veio demonstrar exatamente o contrário.

Ao Pedro Eiras, por todo o apoio e carinho demonstrado ao longo deste último ano. Na verdade, excedeu as minhas melhores expectativas no sentido em que me proporcionou um enorme sentimento de autoconfiança e estímulo para a execução do trabalho.

E, finalmente, a todos os meus amigos, pelo apoio sempre demonstrado, pela sua preocupação e pela disponibilidade. Grata pela dedicação, companheirismo e amizade, fica a tónica na edificação de uma relação interpessoal que jamais esquecerei.

Sensibilizada, deixo o meu profundo agradecimento.

Resumo

A *reconstrução filogenética* é uma área de estudo multidisciplinar onde se juntam a matemática, a biologia e as ciências da computação para permitir construir a *árvore evolutiva* de um conjunto de *objetos biológicos* a partir de relações conhecidas entre eles. Do ponto de vista matemático, a estes *objetos biológicos* vão corresponder os elementos de um conjunto X que estarão em correspondência com os *vértices de grau um*, ditos *folhas*, de um certo *grafo*, dito *X -árvore filogenética*. Nesta dissertação são tratados resultados recentes obtidos por Dress, Huber e Steel, publicados em [4] e [5] que permitem lidar com restrições impostas pelos dados existentes para fazer a reconstrução filogenética. Definem-se e caracterizam-se certos conjuntos de pares de folhas de uma *X -árvore filogenética*, ditos *lassos*, que permitem reconstruir a árvore. Associa-se uma estrutura combinatória, um *matroide*, a uma *X -árvore*. Esse matroide está definido no conjunto de pares de folhas da árvore e mostra-se que as suas *bases* são os *lassos minimais* que permitem reconstruir a árvore.

Palavras-chave: Árvore filogenética, grafo, matroide, lasso, edge-weight lasso, lasso topológico.

Abstract

Phylogenetic reconstruction is a multidisciplinary area in which mathematics, biology and computer science come together aiming to construct the *evolution tree* of a set of *biological objects*, given certain known relations between them. From a mathematical perspective, these *biological objects* correspond to elements of a set X , which in turn correspond to the *degree one vertices*, known as *leaves*, of a certain *graph*, called a *phylogenetic X -tree*. This thesis covers recent results by Dress, Huber, and Steel, published in [4] and [5], which allow for the *phylogenetic reconstruction* under restrictions imposed by existing data. We determine and characterize certain sets of pairs of leaves of a *phylogenetic X -tree*, called *lassos*, which determine aspects of the tree. A combinatorial structure called a *matroid* is associated to a *phylogenetic X -tree*. This matroid is defined on the set of all pairs of leaves of the tree and it is shown that its bases are exactly the *minimal lassos* which allow for the reconstruction of the tree.

Keywords: Phylogenetic tree, graph, matroid, lasso, edge-weight lasso, topological lasso.

Conteúdo

Resumo	v
Abstract	vii
Índice de Figuras	xii
1 Introdução	1
2 Árvores filogenéticas	3
2.1 Conceitos e resultados básicos	3
2.1.1 Grafos	4
2.1.2 Árvores	7
2.1.3 X -árvores e Árvores filogenéticas	12
2.2 Enlaçar uma árvore filogenética	19
3 Matroides	27
3.1 Definição e Propriedades	27
3.2 Circuitos e rank num matroide	31
4 Matroide associado aos lassos de uma árvore filogenética	35
4.1 Definição do matroide	35
4.2 Árvores estrela	46
4.3 Forma recursiva para a construção das bases do matroide	52
4.4 Caracterização da X -árvore pelo seu matroide	60
5 Conclusão	63

Lista de Figuras

2.1	Grafo $G = (V, E)$	4
2.2	Grafos: H , $G[W]$ e $G[F]$	5
2.3	Ciclos do grafo da Figura 2.1.	6
2.4	Grafos Bipartidos.	6
2.5	Grafo conexo G_1 ; grafo desconexo G_2 com duas componentes conexas. . .	7
2.6	Árvore.	7
2.7	Árvores $T \setminus \{e_k\}$: $T \setminus \{fh\}$, $T \setminus \{dg\}$ e $T \setminus \{eh\}$	9
2.8	Ilustração de dois ciclos quando retirada uma aresta.	9
2.9	$T \cup \{gi\}$ e $T \cup \{cd\}$	10
2.10	Árvores geradoras do grafo G da Figura 2.1.	10
2.11	Ilustração da prova.	11
2.12	X -árvore; X -árvore filogenética.	13
2.13	Três X -árvore filogenéticas: T_1 , T_2 e T_3	13
2.14	Árvore com pesos nas arestas.	15
2.15	Árvore com quatro folhas.	16
2.16	Ilustração dos caminhos P_{xy} e P_{zw} ; P_{xz} e P_{yw} ; P_{xw} e P_{yz} , respetivamente. .	16
2.17	Árvore T com os caminhos $P_{v_1v_3}$, $P_{v_3v_5}$, $P_{v_3v_8}$, $P_{v_5v_8}$ e P_{v_1w} a vermelho. . .	17
2.18	X -árvore com pesos nas arestas; função δ associada.	17
2.19	X -árvore com pesos.	20
2.20	Duas X -árvores distintas.	20
2.21	Grafo $\Gamma(\mathcal{L})$ associado a $\mathcal{L} = \{ab, bc, bd, af, ae\}$	21
2.22	X -árvore T	23
2.23	Duas árvores que não são topologicamente equivalentes: T_1 e T_2	24
2.24	T_1 e T_2	24
2.25	X -árvore com conjunto $X = \{a, b, c, d, e, f\}$ e o seu grafo $\Gamma(\mathcal{L})$ associado ao conjunto $\mathcal{L} = \{ab, bc, ac, de, ef, fd\}$	25
2.26	X -árvore com conjunto $X = \{a, b, c, d, e\}$ e o grafo $\Gamma(\mathcal{L})$ associado ao conjunto $\mathcal{L} = \{ab, ac, bc, de\}$	25
2.27	X -árvore com conjunto $X = \{a, b, c, d, e\}$ com diferentes pesos nas arestas. .	26
3.1	Grafo completo com quatro vértices, K_4	30
3.2	Grafo G	31
4.1	X -árvore estrela.	37
4.2	X -árvore estrela.	37
4.3	Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bd, cb, ac\}$	39
4.4	X -árvore quarteto.	40

4.5	Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, cd, ad, bc, bd\}$	42
4.6	X -árvore lagarta.	42
4.7	Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bc, cd, ad, be, ce, bd\}$	44
4.8	X -árvore lagarta.	44
4.9	Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bc, cd, ad, ac, ae, bf, cf, de\}$	45
4.10	X -árvore estrela.	47
4.11	Grafos $\Gamma(\mathcal{L})$ associados aos lassos \mathcal{L} com quatro cordas.	47
4.12	Grafos $\Gamma(\mathcal{L})$ associados aos subconjuntos \mathcal{L} que são bases da X -árvore da Figura 4.10.	48
4.13	Grafo completo K_4	49
4.14	X -árvore estrela.	50
4.15	Duas representações do grafo completo com cinco vértices, K_5	50
4.16	Hipóteses dos ciclos de comprimento três.	51
4.17	Configurações para as bases de $\mathcal{M}(T)$	51
4.18	Configuração para o grafo $\Gamma(\mathcal{L})$ que tem um ciclo de comprimento par.	51
4.19	X -árvore T	52
4.20	X -árvore $T_1 = T/\{f_1\}$	53
4.21	X -árvores: $T_2 = T_1/\{f_2\}$; $T_3 = T_2/\{f_3\}$; $T_4 = T_3/\{f_4\}$	53
4.22	Grafos $\Gamma(\mathcal{L})$, associados às bases do matroide $\mathcal{M}(T)$	56
4.23	Grafos $\Gamma(\mathcal{L})$, associados às bases do matroide $\mathcal{M}(T)$	57
4.24	$\Gamma(B_1)$; $\Gamma(B_2)$	57
4.25	X -árvore com cinco folhas.	58
4.26	$T/\{f, g\}$, árvore estrela com cinco folhas: T^*	58
4.27	Grafos $\Gamma(\mathcal{L})$, associados a um lasso \mathcal{L} que são bases do matroide associado à X -árvore da Figura 4.26.	59
4.28	$T/\{g\}$	59

Capítulo 1

Introdução

The affinities of all the beings of the same class have sometimes been represented by a great tree. As buds give rise by growth to fresh buds, and these if vigorous, branch out and overtop on all sides many a feeble branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications.

Darwin, The Origin of the Species.

Esta dissertação incidirá no estudo de um problema ligado às *árvores filogenéticas*. Estas árvores são estruturas matemáticas que permitem modelar diversos problemas em biologia, em particular permitem representar relações evolutivas entre espécies, como é descrito no texto de Darwin citado acima, ou, por exemplo, entre diferentes indivíduos da mesma espécie.

O grande avanço dos temas da biologia ligados à genética tem contribuído para a evolução do estudo dos modelos matemáticos associados aos problemas por ela levantados, em particular os ligados à chamada *reconstrução filogenética*. Nesta, pretende-se, a partir de relações conhecidas entre *objetos biológicos*, sejam eles diferentes indivíduos da mesma espécie ou espécies diversas, obter informação sobre a sua *árvore evolutiva*. Do ponto de vista matemático, a estes *objetos biológicos* vão corresponder os elementos de um conjunto X que estarão em correspondência com os *vértices de grau um*, ditos *folhas*, de um certo *grafo*, dito *X -árvore filogenética*, como veremos descrito neste texto.

São referidos no texto resultados e algoritmos que permitem fazer essa construção em alguns dos casos em que é conhecida a *distância* entre todos os pares de objetos biológicos associados ao conjunto X referido acima. Contudo, numa das possíveis aplicações em biologia, a sequenciação genética, em geral só é possível determinar a chamada *distância genética* entre uma parte das combinações dos *objetos biológicos* envolvidos, que no modelo matemático serão as folhas da árvore a reconstruir.

Resultados recentes obtidos por Dress, Huber e Steel, publicados em [4] e [5], foram o objetivo central deste estudo e lidam com esta limitação à reconstrução filogenética imposta pelas suas aplicações na biologia.

No primeiro são caracterizados certos conjuntos de pares de folhas de uma X -árvore filogenética, ditos *lassos*, que permitem reconstruir a árvore .

O segundo, utiliza os conceitos e resultados do primeiro e associa uma estrutura combinatória, um *matroide*, a uma X -árvore. Esse matroide está associado ao conjunto de pares de folhas da árvore e as suas *bases* são os *lassos minimais* que permitem reconstruir a árvore.

Estruturamos o resto desta dissertação em três capítulos, organizados do modo que a seguir descrevemos.

Na primeira parte do segundo capítulo introduzimos conceitos e resultados básicos de teoria de grafos, em particular, a noção de X -árvore *filogenética* e a de *distância* entre as folhas de uma tal árvore. Os exemplos e resultados apresentados pretendem permitir a compreensão de um dos resultados clássicos e fundamentais da filogenética, o *Tree Metric Theorem*.

Na segunda parte, com o intuito de investigar quais dos subconjuntos de $\binom{X}{2}$ são suficientes para determinar a X -árvore introduzimos o conceito de *lasso*, caracterizando os *lassos* que determinam a topologia da árvore (*lassos topológicos*), os pesos entre todos os pares de folhas (*edge-weight lasso*) ou ambos (*strong lasso*).

No terceiro capítulo, para além de ser dada a definição de *matroide*, são apresentados exemplos relacionados com os resultados objeto do nosso estudo, como o matroide gráfico e o matroide vetorial. São também vistas diferentes caracterizações desta estrutura combinatória, usando os conjuntos de *bases*, dos *circuitos* ou a *função rank* que, além de serem necessárias para a compreensão dos resultados principais, permitem melhor compreender esta estrutura.

Dedicamos o último capítulo ao estudo do matroide associado a uma X -árvore filogenética. Ele é definido no conjunto dos pares de elementos do conjunto X , usando a *função rank* e são vistos com detalhe exemplos desse matroide para algumas X -árvores.

Utilizamos resultados referidos nos capítulos anteriores que nos permitem mostrar que um *edge-weight lasso* é um gerador deste matroide e um *edge-weight lasso* de cardinalidade mínima é uma base. Nas secções seguintes usamos um dos resultados apresentado em [5] que caracteriza este matroide no caso das *árvores estrela* para fazer alguns exemplos desse matroide e descrever uma construção recursiva para a construção do conjunto das bases do matroide de uma X -árvore a partir do mesmo conjunto para as *árvores estrela*.

Concluimos a tese com um dos resultados principais de [5] que mostra que este matroide caracteriza a X -árvore.

Capítulo 2

Árvores filogenéticas

Neste capítulo vamos estudar noções básicas da teoria de grafos. Iremos focar-nos num tipo de grafos especial, as *árvores*, com o intuito de tornar compreensíveis alguns conceitos e resultados sobre *árvores filogenéticas*, objeto central desta dissertação.

Depois de introduzidas as definições, notações e resultados necessários ao enquadramento destes grafos, explicamos, na segunda parte do capítulo, a noção de *lasso* de uma árvore filogenética, tratando alguns exemplos e abordando alguns resultados que envolve este conceito introduzido em [4]; alguns dos conceitos e resultados aqui introduzidos serão utilizados posteriormente.

2.1 Conceitos e resultados básicos

A teoria dos grafos, na história da matemática, teve origem no século *XVIII*. Esta teoria não se aplica só na matemática, mas também nas áreas de física, química, biologia, ciência da computação e economia.

Um grafo é uma noção simples, abstrata e intuitiva, com a qual se representam relações, binárias, entre objetos. São usualmente representados por uma figura com pontos, denominados por *vértices*, que correspondem aos objetos; dois desses pontos são unidos por uma linha, denominada por *aresta*, sempre que esse par de objetos pertence à relação .

O mais famoso problema e também aquele ao qual se atribui usualmente a origem da teoria de grafos é o problema das pontes de Königsberg. Nesta cidade, sete pontes cruzam o rio Pregel, estabelecendo ligações entre duas ilhas e entre as ilhas e as margens opostas do rio. Dizia-se que os habitantes dessa cidade tentavam efetuar um percurso passando por todas as pontes, mas apenas uma vez em cada, voltando no final ao local de partida. Os habitantes pensavam que não era possível, mas não sabiam porquê. Euler, em 1735, provou que tal percurso não existia e deu uma caracterização geral dos *grafos* em que existem percursos deste tipo. Os vértices do grafo associado a este problema correspondem às quatro regiões da cidade e uma aresta é definida para cada ponte que liga duas das regiões.

Os problemas de percurso, como o referido ou como, por exemplo, os associados a redes de transportes, têm modelação por grafos quase natural, tanto mais quanto se

permitir a existência de várias arestas a ligar dois vértices, como em Königsberg, ou se for introduzida orientação nas arestas como interessará em rede de transportes.

A definição que daremos aqui de grafo é de algum modo a mais simples porque também é a conveniente para o problema que iremos tratar.

2.1.1 Grafos

Definição 1. Dado um conjunto A , definimos o conjunto $\binom{A}{2}$ como o subconjunto das partes de A com exatamente dois elementos, i.e., o conjunto dos pares de objetos de A .

Definição 2. Um *grafo* G é definido pelo par ordenado de conjuntos (V, E) , onde $V = V(G) = \{v_1, v_2, \dots, v_n\}$ é um conjunto não vazio e finito, cujos elementos se designam por *vértices* e $E = E(G) = \{e_1, \dots, e_m\}$ é um conjunto de pares não ordenados de vértices, isto é, $E \subseteq \binom{V}{2}$, aos quais chamamos *arestas*.

Exemplo 1. O diagrama abaixo representa o grafo G definido por:

$$\begin{aligned} V(G) &= \{a, b, c, d, e, f, g, h, i\} \\ E(G) &= \{\{a, b\}, \{a, d\}, \{b, c\}, \{b, e\}, \{b, f\}, \{c, d\}, \{c, f\}, \{d, g\}, \{e, h\}, \\ &\quad \{f, g\}, \{f, h\}, \{f, i\}, \{g, i\}, \{h, i\}\} \end{aligned}$$

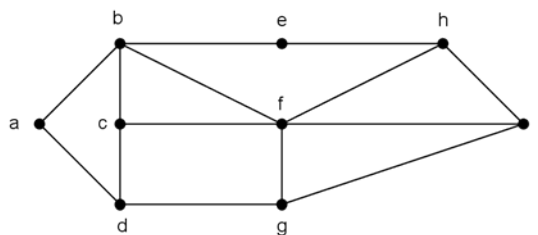


Figura 2.1: Grafo $G = (V, E)$

Dois vértices u e v de um grafo $G = (V, E)$ dizem-se *adjacentes* se $\{u, v\} \in E$, ou seja, se $\{u, v\}$ é uma aresta do grafo, vamos denotar por uv essa aresta. Esses vértices dizem-se *extremos* da aresta uv . A aresta uv , diz-se *incidente* em u e em v e estes vértices também se dizem incidentes na aresta.

Definição 3. O *grau* de um vértice v em G é o número de arestas $e \in E(G)$ incidentes nele. Vamos denotar o grau do vértice v por $\deg(v)$. Por exemplo, na Figura 2.1, $\deg(a) = 2$ e $\deg(f) = 5$.

Definição 4. Se $G = (V, E)$ é um grafo:

- qualquer grafo $H = (W, F)$ tal que $W \subseteq V$ e $F \subseteq E$ diz-se um *subgrafo* do grafo $G = (V, E)$.
- um subgrafo $H = (W, F)$ do grafo $G = (V, E)$ diz-se um *subgrafo gerador* se $W = V$.

- se $W \subseteq V$ é um conjunto de vértices de G , o *subgrafo induzido* por W em G é o grafo $G[W] = (W, E_W)$, onde E_W é o conjunto de arestas de G com ambas as extremidades em W .
- se $F \subseteq E$ é um conjunto de arestas de G , o subgrafo induzido por F em G é o grafo $G[F] = (V_F, F)$, onde V_F é o conjunto de vértices de G incidentes nas arestas de F .

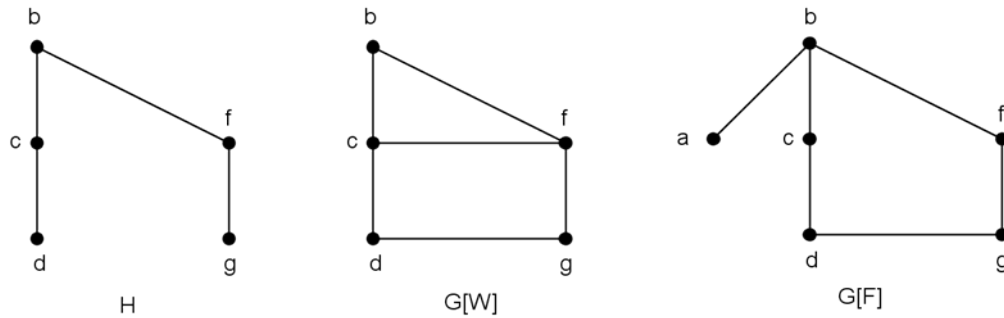


Figura 2.2: Grafos: H , $G[W]$ e $G[F]$.

Os três grafos representados na Figura 2.2 são subgrafos de $G = (V, E)$, representado na Figura 2.1.

O grafo $H = (W, F)$ é um subgrafo de G com, $W = \{b, c, d, f, g\} \subseteq V$ o seu conjunto de vértices e $F = \{bc, cd, bf, fg\} \subseteq E$ o conjunto de arestas.

O grafo $G[W] = (W, E_W)$ é o subgrafo induzido por G no mesmo conjunto $W = \{b, c, d, f, g\} \subseteq V$; E_W é o conjunto de arestas de G com ambas as extremidades em W , $E_W = \{bc, bf, cf, cd, fg, dg\}$.

O grafo $G[F] = (V_F, F)$ é o subgrafo induzido por G no conjunto de arestas $F = \{ab, bc, bf, cd, fg, dg\}$; o seu conjunto de vértices é $V_F = \{a, b, c, d, f, g\}$ é o conjunto de vértices de G incidentes nas arestas do conjunto F .

Definição 5. Um *caminho* de um grafo G é uma sequência de vértices todos distintos, em que dois quaisquer vértices consecutivos são adjacentes. Se o vértice inicial do caminho for u e o vértice final for v diz-se que é o caminho de u a v ou entre u e v .

Um grafo $P_k = (V, E)$, onde $V = \{v_0, v_1, \dots, v_k\}$ e $E = \{v_0v_1, \dots, v_{k-1}v_k\}$ diz-se um *caminho* de comprimento k , um caminho em G é um subgrafo de G .

Na Figura 2.1, $abfh$, $adcfh$, $adgfh$ e $adgfh$ são caminhos entre a e h . Estes quatro caminhos são distintos; $adcfh$ e $adgfh$ têm o mesmo *comprimento*, ou seja, o mesmo número de arestas; $abfh$ é um caminho de *comprimento mínimo*, pois contém o menor número de arestas necessárias para ligar a a h em G .

Definição 6. Um *ciclo* de um grafo é uma sequência de vértices, em que o primeiro e o último elemento são iguais e todos os outros são distintos entre si e distintos deste; além disso, dois quaisquer vértices consecutivos desta sequência são adjacentes e ela tem pelo menos três vértices distintos.

Observe-se que num ciclo, o número de vértices e o de arestas coincidem, e esse número diz-se o *comprimento do ciclo*; além disso, no ciclo, qualquer vértice tem grau dois e o seu comprimento é ≥ 3 .

Por exemplo, na Figura 2.3 estão representados a vermelho os ciclos $abcd$ de comprimento quatro e o ciclo $figf$ de comprimento três.

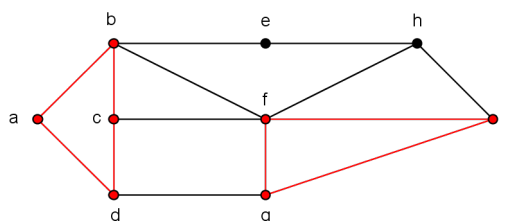


Figura 2.3: Ciclos do grafo da Figura 2.1.

Definição 7. Um grafo $G = (V, E)$ diz-se *bipartido*, se o seu conjunto de vértices pode ser subdividido em dois subconjuntos, V_1 e V_2 ; $V = V_1 \cup V_2$ e $V_1 \cap V_2 = \emptyset$, tais que não existem arestas entre os elementos do mesmo subconjunto de vértices.

Observe-se que decorre da definição que toda a aresta do grafo bipartido G une um vértice de V_1 a um vértice de V_2 . Em particular, um grafo *vazio*, $G = (V, \emptyset)$, é bipartido.

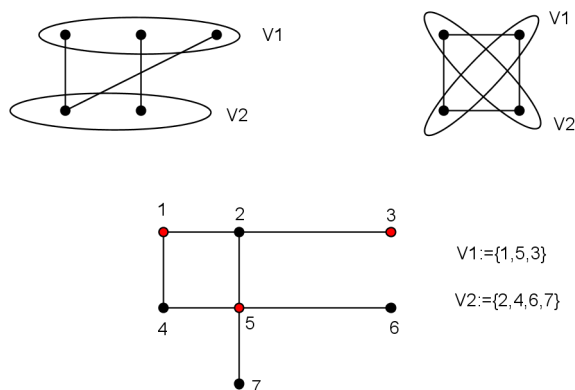


Figura 2.4: Grafos Bipartidos.

Teorema 2.1. *Um grafo é bipartido se e só se não possui ciclos de comprimento ímpar.*

A prova do Teorema 2.1 pode ser vista na página 16 de [10].

Observe-se que, em consequência do Teorema 2.1, os ciclos de comprimento par e os grafos sem ciclos, ditos *acíclicos*, são grafos bipartidos.

Definição 8. Um grafo $G = (V, E)$ diz-se *conexo* se dados dois quaisquer dos seus vértices existe um caminho entre eles em G .

Um grafo $G = (V, E)$ que não é conexo diz-se *desconexo* e pode sempre ser decomposto como união disjunta de subgrafos conexos maximais, ditas as componentes conexas de G .

Mais precisamente, $G = G[V_1] \dot{\cup} G[V_2] \dot{\cup} \dots \dot{\cup} G[V_k]$, onde $V = V_1 \dot{\cup} V_2 \dot{\cup} \dots \dot{\cup} V_k$, $G[V_i]$ é um grafo conexo maximal, para todo $i = 1, 2, \dots, k$ (i.e., qualquer subgrafo de G contendo estritamente algum destes $G[V_i]$ é desconexo).

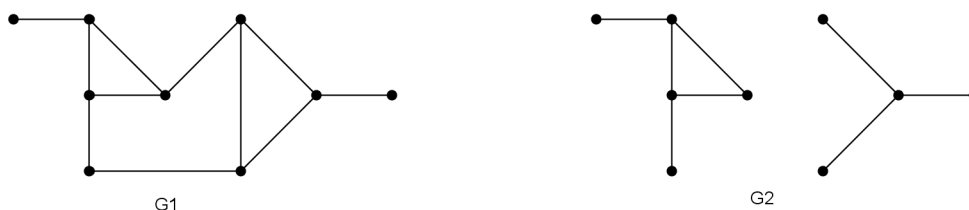


Figura 2.5: Grafo conexo G_1 ; grafo desconexo G_2 com duas componentes conexas.

2.1.2 Árvores

De entre os grafos conexos, as *árvores* são, de um certo ponto de vista, os mais simples mas são também aqueles que surgem num grande número de aplicações em áreas muito diversas. Em particular, elas modelam diversos problemas ligados à genética, entre eles os que estão ligados ao trabalho sobre o qual nos debruçamos nesta dissertação.

Definição 9. Uma *árvore* é um grafo conexo que não contém ciclos.

Observe-se que se um grafo é *acíclico* todas as suas componentes conexas são árvores e utiliza-se a designação de *floresta* para estes grafos.

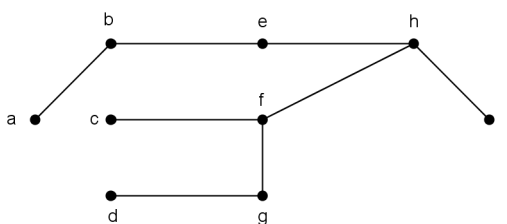


Figura 2.6: Árvore.

Numa árvore, chama-se *folha* a qualquer vértice de grau um; na árvore da Figura 2.6 o conjunto de folhas é $\{a, c, d, i\}$. Os vértices com grau diferente de um são designados por *vértices interiores*; na árvore da Figura 2.6 o conjunto dos vértices interiores é $\{b, e, h, f, g\}$.

Lema 1. Existe exatamente um caminho entre cada par de vértices de uma árvore T .

Demonstração. Seja T uma árvore e sejam x e y dois vértices de T .

Como T é conexo, existe em T um caminho entre x e y ; seja $xx_1x_2\dots x_ky$ esse caminho. Resta provar que este caminho é único.

Admitamos que existe outro caminho $xy_1y_2\dots y_ly$ e consideremos a justaposição do primeiro com o inverso do segundo, isto é, a sequência de vértices

$$xx_1x_2\dots x_kyy_ly_{l-1}\dots y_1x.$$

É claro que quaisquer dois vértices consecutivos desta nova sequência são adjacentes.

Se os vértices diferentes de x desta sequência forem todos distintos, ela forma um ciclo do grafo, o que contradiz o facto de T ser uma árvore.

Em geral, eles não são necessariamente todos distintos, mas é fácil ver que vai sempre existir uma subsequência destes que forma um ciclo. \square

Em consequência do Lema 1, dados dois vértices a e b de uma árvore, fica bem definido o caminho entre eles, que será denotado por P_{ab} ; ele é o único caminho na árvore que liga estes dois vértices.

Vamos denotar por E_{ab} o conjunto das arestas de P_{ab} .

Por exemplo, na Figura 2.6, $P_{ai} = abehi$ e $E_T(a|i) = \{ab, be, eh, hi\}$.

Observações:

- Seja $G = (V, E)$ um grafo e e uma aresta do grafo G , denota-se por $G \setminus \{e\}$ o grafo obtido a partir de G quando se retira a aresta e , i.e., o subgrafo de G definido por $V(G \setminus \{e\}) = V(G)$ e $E(G \setminus \{e\}) = E(G) \setminus \{e\} \subseteq E_G$.
- Quando a uma árvore T , retirarmos uma aresta $e \in E(T)$ obtemos um grafo, $T \setminus e$, que é na mesma acíclico, claro, mas desconexo já que ao retirar a aresta $e = uv$ os vértices u e v não podem permanecer ligados por um caminho em $T \setminus e$ porque esse caminho juntamente com a aresta e formaria um ciclo de T . Além disso, pode-se provar que $T \setminus e$ tem exatamente duas componentes conexas. Cada uma destas componente é, então, um grafo conexo e acíclico, ou seja, uma árvore.

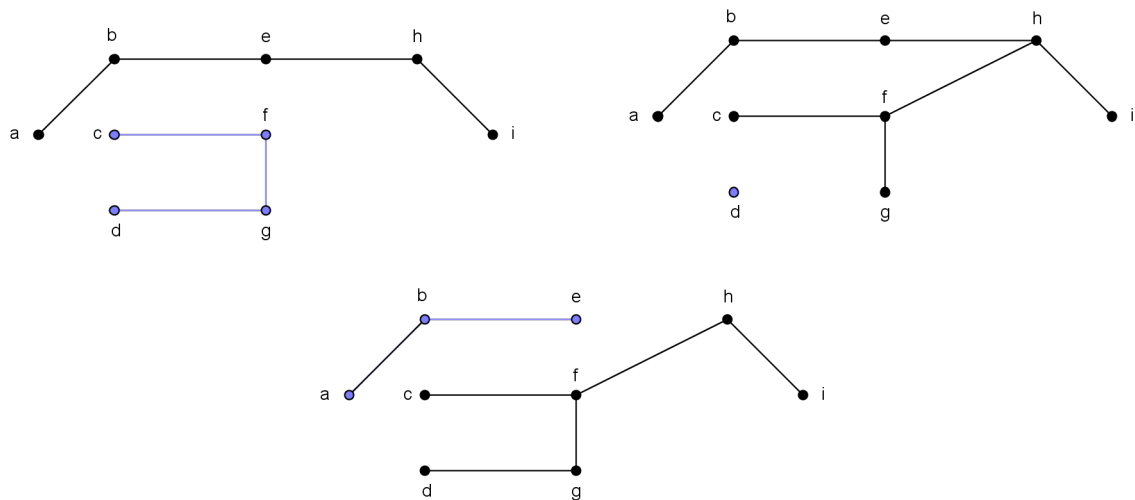


Figura 2.7: Árvores $T \setminus \{e_k\}$: $T \setminus \{fh\}$, $T \setminus \{dg\}$ e $T \setminus \{eh\}$.

Na Figura 2.7 estão representadas três florestas obtidas da árvore da Figura 2.6, quando retirada uma das suas arestas. As duas componentes conexas obtidas em cada um dos casos estão representadas uma a azul e outra a preto.

- Se C é um ciclo e $e = xy \in E(C)$, $C \setminus \{e\}$ é um caminho entre x e y e, portanto, um grafo conexo.

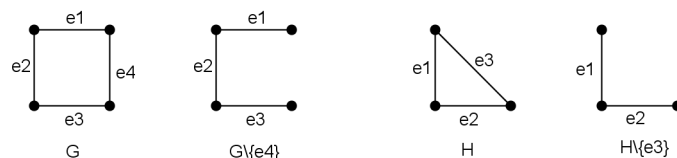


Figura 2.8: Ilustração de dois ciclos quando retirada uma aresta.

Teorema 2.2. O número de arestas de uma árvore T , dito cardinalidade da árvore e notado por $|T|$, só depende do número n de vértices de T : $|T| = n - 1$.

Demonstração. Nos casos $n = 1$ e $n = 2$ o resultado verifica-se.

Para provarmos o resultado por indução consideremos T uma árvore qualquer com n vértices, $n \geq 3$, e retiremos-lhe uma aresta.

Como observámos acima, se retirarmos uma aresta e a uma árvore T , $T \setminus e$ é uma floresta constituída por duas árvores, T' e T'' ; sejam n' e n'' , respetivamente, o número de vértices dessas árvores e seja n o número de vértices de T . É claro que $n = n' + n''$.

Por hipótese de indução, o número de arestas de T' é $n' - 1$ e o número de arestas de T'' é $n'' - 1$.

Assim o número de arestas da árvore T é:

$$(n'' - 1) + (n'' - 1) + 1 = n - 1$$

como queríamos demonstrar. \square

Se ao invés de retirarmos uma aresta à árvore T lhe acrescentarmos uma aresta $e = xy$ ligando dois dos seus vértices não adjacentes, forma-se um ciclo deste novo grafo, $T \cup \{e\}$. Esse ciclo é obtido pela adição dessa aresta ao caminho de T entre x e y ; e ele é o único ciclo do grafo, $T \cup \{e\}$ porque, se não fosse único, como qualquer ciclo teria de conter e , ao retirar a aresta xy existiriam dois caminhos distintos entre x e y em T , o que não pode acontecer, pois T é uma árvore.

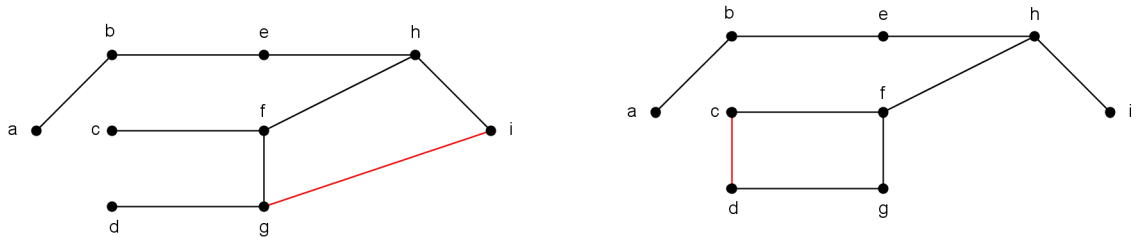


Figura 2.9: $T \cup \{gi\}$ e $T \cup \{cd\}$.

Na Figura 2.9 estão representados os dois grafos obtidos da árvore da Figura 2.6, quando adicionada as arestas gi e cd , respectivamente. A aresta adicionada está representada a vermelho.

Definição 10. Uma *árvore geradora* de um grafo G é qualquer árvore que é um subgrafo gerador de G .

Na Figura 2.10, os subgrafos indicados em linhas vermelhas são duas árvores geradoras do grafo G da Figura 2.1.

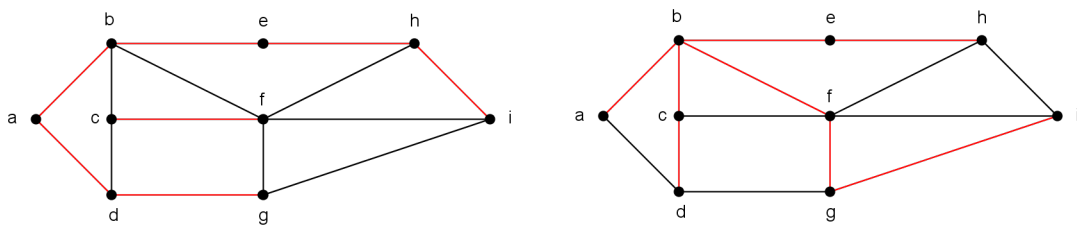


Figura 2.10: Árvores geradoras do grafo G da Figura 2.1.

Observações:

- Decorre do Teorema 2.2 que a *cardinalidade de uma árvore geradora* de um grafo G com n vértices é igual a $n - 1$ e, conseqüentemente, que todas as árvores geradoras do grafo G têm a mesma cardinalidade.

- Qualquer grafo conexo G tem, no mínimo, uma árvore geradora.

Demonstração. Seja então G um grafo conexo. Se G for um grafo acíclico então G é uma árvore; neste caso há uma só árvore geradora do grafo G que é o próprio grafo.

Caso contrário, G tem pelo menos um ciclo: consideremos uma aresta e de G tal que e pertença a um ciclo do grafo G e retiremos essa aresta a G . Obtemos assim o grafo $G \setminus \{e\}$ que é um subgrafo de G , gerador porque tem o mesmo conjunto de vértices e conexo porque a aresta retirada pertencia a um ciclo.

Se $G \setminus \{e\}$ não tiver ciclos é uma árvore geradora de G . Caso contrário, retiramos outra aresta que pertença a um ciclo do grafo $G \setminus \{e\}$ e assim sucessivamente. Como os subgrafos obtidos a cada passo são sempre geradores e conexos e têm cada um deles menos ciclos que o anterior, ao fim de um número finito de passos o subgrafo obtido será também acíclico e, portanto, uma árvore geradora do grafo G .

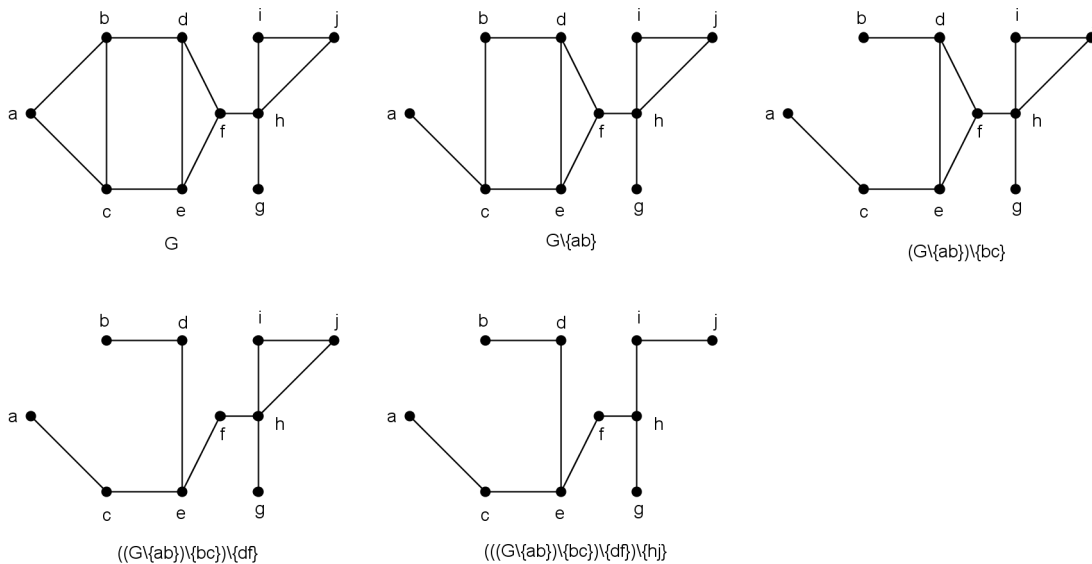


Figura 2.11: Ilustração da prova.

□

Lema 2. Seja $G = (V, E)$ um grafo conexo com n vértices, e $F \subseteq E$ um conjunto acíclico de arestas de G , i.e., um conjunto $F \subseteq E$ tal que o grafo $G[F] = (V_F, F)$ induzido por F em G é acíclico.

Então, F é o conjunto das arestas de uma árvore geradora se e só se F é um conjunto maximal para a inclusão com esta propriedade, i.e., se para todo X tal que $F \subset X \subseteq E$, $G[X]$ não é acíclico.

Demonstração. Se F é o conjunto das arestas de uma árvore geradora, $T = G[F]$, então $V_F = V$ e $|F| = |V| - 1$.

Assim, para todo X tal que $F \subset X \subseteq E$, como T é subgrafo de $G[X]$, $G[X]$ é conexo e $V_X = V$ portanto se $G[X]$ fosse acíclico seria uma árvore geradora com pelo menos $|V|$ arestas, contradizendo o Teorema 2.2.

Reciprocamente, se o conjunto F é acíclico maximal, então o grafo $G[F]$ tem de ser conexo e $V_F = V$, i.e., uma árvore geradora de G ; caso contrário, como G é conexo, seria possível acrescentar pelo menos uma aresta $e \in E$ ao conjunto F sem formar ciclos no grafo induzido pelo novo conjunto $F \cup \{e\}$.

□

2.1.3 X -árvores e Árvores filogenéticas

Devido à sua estrutura, as árvores são uma forma de representar a evolução das espécies e a forma como se relacionam entre si. Em particular, as árvores, denominadas por árvores filogenéticas, são um bom modelo para a representação das relações evolutivas entre várias espécies ou, por exemplo, entre diferentes indivíduos da mesma espécie.

Nestas árvores, os vértices de grau 1 são etiquetados por elementos de um conjunto X , conjunto que é inicialmente fixado e usualmente representa o conjunto das espécies ou indivíduos cuja evolução se pretende modelar. Os vértices que ficam sem etiqueta, vértices “intermédios”, terão obrigatoriamente grau maior do que dois, com exceção feita, por vezes, a um vértice destacado, dito a *raíz*, que geralmente representa um ancestral comum aos elementos de X . Nestes casos, o caminho que liga a raíz a um vértice de etiqueta $x_0, x_0 \in X$, ordena os antepassados de x_0 até esse ancestral; o primeiro vértice comum aos caminhos dos vértices de etiquetas $x, y \in X$ representa o mais “recente antepassado comum a ambos. Nesta árvore é também usual atribuir um peso às arestas que pode, por exemplo, representar o tempo de evolução entre as espécies que a aresta liga.

Os resultados de que falaremos nesta secção são válidos num contexto um pouco mais geral das chamadas X -árvores, que definimos abaixo, de acordo com a opção feita no livro [3] onde este tema é tratado com detalhe.

No resto do trabalho trataremos um caso particular destas árvores, definidas abaixo como X -árvores filogenéticas

Definição 11. Seja X um conjunto; uma X -árvore é um par ordenado $\Gamma = (T, \phi)$, em que $T = (V, E)$ é uma árvore e $\phi : X \rightarrow V$ é uma função com a propriedade de todos os vértices de T de grau ≤ 2 pertencerem à sua imagem.

Definição 12. Uma X -árvore filogenética é uma X -árvore $\Gamma = (T, \phi)$, em que ϕ é uma bijeção de X no conjunto das folhas de T .

Decorre da definição que uma X -árvore filogenética não pode ter vértices de grau 2.

Se além disso, todos os vértices interiores de T tiverem grau três, então Γ é uma X -árvore filogenética binária.

Na Figura 2.12 estão representadas uma X -árvore e uma X árvore filogenética com os conjuntos de etiquetas:

$$X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}; X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

respectivamente.

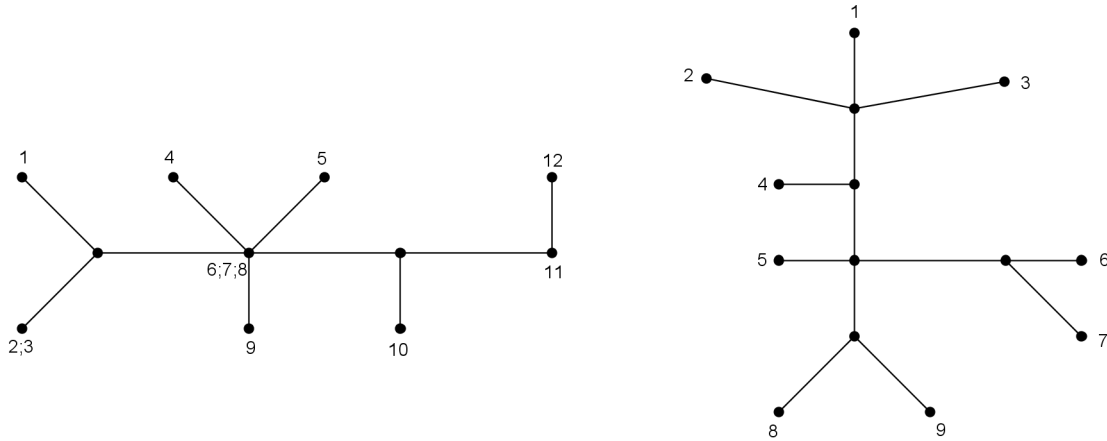


Figura 2.12: X -árvore; X -árvore filogenética.

No caso das X -árvores filogenéticas, é usual, e também o faremos ao longo deste trabalho, identificar o conjunto X das etiquetas com o conjunto das folhas de T .

Definição 13. Duas X -árvores filogenéticas, $T_1 = (V_1, E_1)$ e $T_2 = (V_2, E_2)$, são *isomorfas* se existe uma função $\phi : V_1 \rightarrow V_2$, tal que:

1. ϕ é bijetiva;
2. $\phi|_X = Id$;
3. $ab \in E(T_1) \Leftrightarrow \phi(a)\phi(b) \in E(T_2)$.

Se T_1 é isomorfa a T_2 , ou equivalentemente, se as duas X -árvores são *topologicamente equivalentes*, escrevemos $T_1 \stackrel{X}{\cong} T_2$.

Exemplo 2. Na Figura 2.13 estão representadas três X -árvore filgenéticas, T_1 , T_2 e T_3 . A X -árvore T_1 é isomorfa à X -árvore T_2 , $T_1 \stackrel{X}{\cong} T_2$, mas não é isomorfa à X -árvore T_3 , $T_1 \not\stackrel{X}{\cong} T_3$.

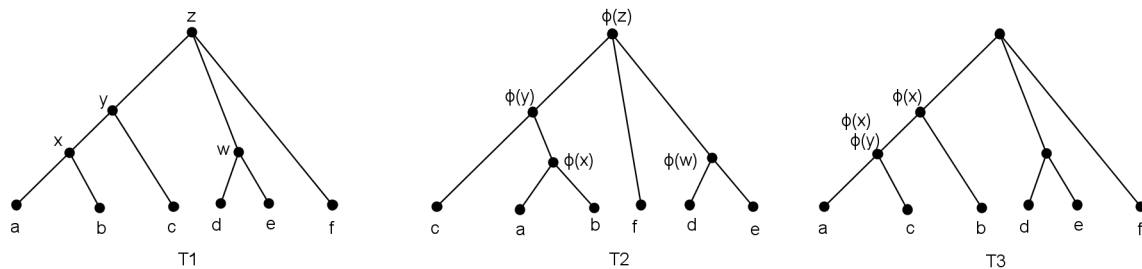


Figura 2.13: Três X -árvore filogenéticas: T_1 , T_2 e T_3 .

As duas X -árvores, T_1 e T_2 , são isomorfas pois a função $\phi : V(T_1) \rightarrow V(T_2)$ representada na Figura 2.13 respeita os três pontos da Definição 13, portanto é um isomorfismo.

Vamos verificar que as duas X -árvores, T_1 e T_3 , não são isomorfas. Admitamos que existe um isomorfismo $\phi : V(T_1) \rightarrow V(T_2)$ entre estas duas árvores, representado na Figura 2.13. Observando a X -árvore T_1 , sabemos que, a aresta ax tem que ter por imagem uma aresta com uma extremidade em a e outra em $\phi(x)$, com $\phi(x)$ o vértice correspondente à imagem de x , na X -árvore T_3 , ilustrado na Figura 2.13. Do mesmo modo, a aresta bx tem que ter por imagem uma aresta com uma extremidade b a outra em $\phi(y)$, com $\phi(y)$ o vértice correspondente à imagem de y , na X -árvore T_3 .

Assim, $\phi(x) = \phi(y)$, o que não pode acontecer pois a função é bijetiva, logo $T_1 \not\cong T_3$.

Como já referimos, nas aplicações à biologia pode ser importante atribuir um peso às arestas da árvore que se pretende ser um modelo dos problemas a tratar. Vamos ver que esses pesos permitem determinar *distâncias* entre as folhas e que reciprocamente, em certas condições é possível recuperar a árvore a partir dessas distâncias.

Vamos considerar $T = (V, E)$ uma árvore com conjunto de vértices V , e com conjunto de arestas $E \subseteq \binom{V}{2}$. Seja w , uma função definida no conjunto das arestas, que associa a cada aresta e um número real positivo, $w(e)$, dito o seu *peso*:

$$\begin{aligned} E: w &\longrightarrow \mathbb{R}^+ \\ e &\longmapsto w(e) \end{aligned}$$

Esta função, denominada por *função peso*, permite definir uma outra função, D_w , no conjunto dos pares ordenados de folhas de T . Esta atribui a cada par, o *peso do único caminho* existente entre os vértices do par na árvore T , do seguinte modo:

$$\begin{aligned} D_w = D_{(T,w)} : X \times X &\longrightarrow \mathbb{R}_{\geq 0} : \\ (x, y) &\longmapsto D_w(x, y) := w_+(E_T(x|y)) \end{aligned}$$

onde

- X é o conjunto das folhas de T ;
- $E_T(x|y)$ denota, para quaisquer duas folhas $x, y \in X$, o conjunto das arestas de T do caminho que liga x a y ;
- Em geral, pode estender-se a notação, fazendo para $\eta \subseteq E$, $w_+(\eta) = \sum_{e \in \eta} w(e)$ e $w_+(\emptyset) = 0$.

Esta função, D_w , verifica as propriedades que caracterizam qualquer métrica, indicadas de seguida.

- D_w é positiva, isto é, $D_w(x, y) \geq 0$ para quaisquer $x, y \in V(T)$ e $D_w(x, y) = 0 \Leftrightarrow x = y$.

Sejam $x = v_0$ e $y = v_k$ dois vértices da árvore T . Como foi definido anteriormente, o peso de todas as arestas de T é positivo, logo é claro que:

$$D(x, y) = w(E_T(x, y)) = w(e_0 = \{v_0, v_1\}) + w(e_1 = \{v_1, v_2\}) + \dots + w(e_k = \{v_{k-1}, v_k\}) \geq 0.$$

Se $x \neq y$ então $E_T(x, y) \neq \emptyset$ e $w_+(E_T(x, y)) > 0$ porque $w(e) > 0$, $\forall e \in E$.

Se $x = y$ então $E_T(x, y) = \emptyset$ e $w_+(E_T(x, y)) = 0$.

- D_w é simétrica, isto é, $D_w(x, y) = D_w(y, x)$.

Como, sempre que $v_0v_1\dots v_k$ é um caminho de x a y então $v_k\dots v_1v_0$ é o caminho de y a x , e, como nas árvores esse caminho é único então:

$$E_T(x, y) = E_T(y, x) \text{ e } D_w(x, y) = D_w(y, x).$$

- D_w verifica a desigualdade triangular, isto é, $D_w(x, z) \leq D_w(x, y) + D_w(y, z)$ para quaisquer $x, y, z \in X$.

Seja $Q = P_{xy}P_{yz}$ a justaposição do caminho entre x e y com o caminho entre y e z .

Então Q é uma sequência de vértices, com vértice inicial x , vértice final z e tal que, vértices consecutivos são adjacentes. Resulta então, que o caminho entre x e z é um subconjunto de Q . Em particular:

$$E_T(x, z) \subseteq E_T(x, y) \cup E_T(y, z)$$

e portanto

$$w_T(E_T(x, z)) \leq w_T(E_T(x, y)) + w_T(E_T(y, z)).$$

Logo, D_w é uma métrica.

Exemplo 3. Vamos observar a X -árvore representada na Figura 2.14, com pesos nas suas arestas.

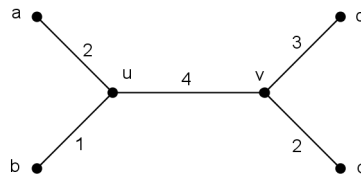


Figura 2.14: Árvore com pesos nas arestas.

O caminho entre a a c é dado por: $E_T(a|c) = \{\{a, u\}, \{u, v\}, \{v, c\}\}$ e assim $D_w(a, c) = 2 + 4 + 3 = 9$.

A partir da X -árvore constrói-se a matriz de distâncias seguinte:

$$\begin{matrix} & \begin{matrix} a & b & c & d \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \end{matrix} & \begin{pmatrix} 0 & 3 & 9 & 8 \\ & 0 & 8 & 7 \\ & & 0 & 5 \\ & & & 0 \end{pmatrix} \end{matrix}$$

Com o peso dos caminhos entre todos os pares de folhas da árvore, calculamos as três somas:

$$\begin{aligned} D(a, b) + D(c, d) &= 3 + 5 = 8 \\ D(a, c) + D(b, d) &= 9 + 7 = 16 \\ D(a, d) + D(b, c) &= 8 + 8 = 16. \end{aligned}$$

Observamos que duas das somas são iguais e o seu valor é superior à terceira soma. É fácil ver que esta relação não depende do valor dos pesos das arestas. Em geral, considere-se a X -árvore representada na Figura 2.15, uma X -árvore sem pesos nas arestas.

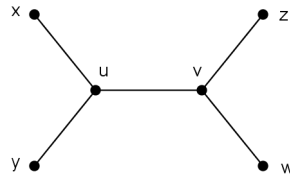


Figura 2.15: Árvore com quatro folhas.

Esta X -árvore, uma X -árvore filogenética binária com quatro folhas, vai ser denominada por *árvore quarteto*.

Tendo em atenção o Exemplo 3 e a Figura 2.16 é fácil verificar que para o caso simples da árvore representada na Figura 2.15 a seguinte condição é satisfeita:

$$D(x, y) + D(z, w) \leq D(x, z) + D(y, w) = D(x, w) + D(y, z)$$

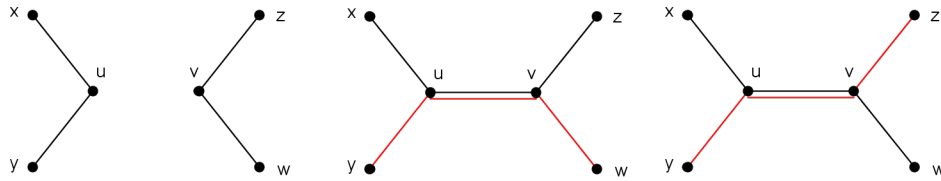


Figura 2.16: Ilustração dos caminhos P_{xy} e P_{zw} ; P_{xz} e P_{yw} ; P_{xw} e P_{yz} , respetivamente.

Este é um exemplo em que a desigualdade da Definição 14 é verificada.

Definição 14 (Condição dos quatro pontos). Numa árvore T , onde está definida uma função $D : X \times X \rightarrow \mathbb{R}_{\geq 0}$ dizemos que a *condição dos quatro pontos* é satisfeita, se para quaisquer folhas $a, b, c, d \in V$, não necessariamente distintas, duas das seguintes somas

$$\begin{aligned} D(a, b) + D(c, d) \\ D(a, c) + D(b, d) \\ D(a, d) + D(b, c) \end{aligned}$$

são iguais e o seu valor é maior ou igual à terceira soma.

Lema 3. Seja $T = (V, E)$ uma X -árvore e $w : E \rightarrow \mathbb{R}^+$ uma função peso, então T verifica a condição dos quatro pontos para a métrica D_w .

A prova resulta de observar que para quaisquer quatro folhas, como no caso de v_1, v_3, v_5, v_8 na Figura 2.17, os caminhos envolvidos no cálculo das distâncias para a condição dos quatro pontos podem ser vistos como os do caso simples representado na Figura 2.15, considerando $x = v_1, y = v_3, z = v_5$ e $w = v_8$.

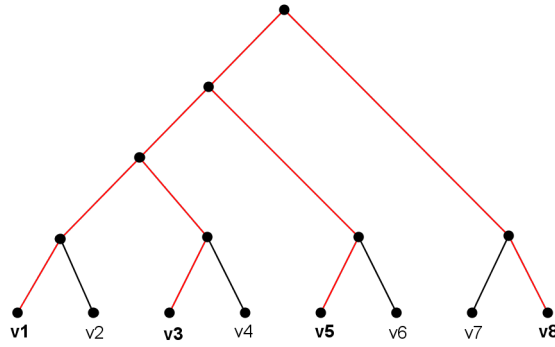


Figura 2.17: Árvore T com os caminhos $P_{v_1v_3}$, $P_{v_3v_5}$, $P_{v_3v_8}$, $P_{v_5v_8}$ e $P_{v_1v_8}$ a vermelho.

Exemplo 4. Na Figura 2.18 está representada uma árvore com uma métrica D_w determinada pela função peso representada nas arestas da figura.

A métrica D_w definida em $X \times X$, $X = \{a, b, c, d, e\}$, pode ser representada pela matriz 5×5 , $\delta = (\delta_{ij})$, dita *matriz de distâncias*, $\delta_{ij} = D_w(i, j)$ com $(i, j) \in X \times X$.

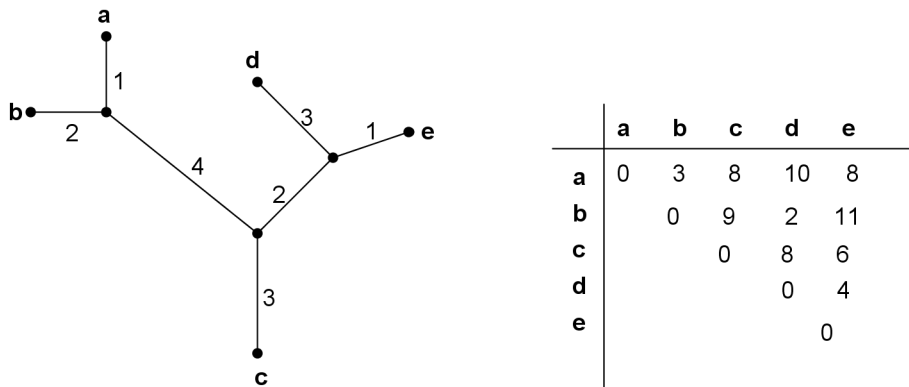


Figura 2.18: X -árvore com pesos nas arestas; função δ associada.

Em geral, dada uma X -árvore T , com uma função peso positiva nas arestas, podemos associar-lhe uma função:

$$\delta : X \times X \rightarrow \mathbb{R}^+$$

com $\delta(x, y) = D_w(x, y) \forall x, y \in X$.

Esta função pode ser representada por uma matriz $n \times n$, $n = |X|$, dita *matriz de distâncias* $\delta = (\delta_{ij})$, em que cada entrada δ_{ij} é definida como o peso do caminho entre o vértice i e o vértice j na árvore T .

Como $\delta(x, x) = 0$ e $\delta(x, y) = \delta(y, x) \forall x, y \in X$. Esta matriz é simétrica, com zeros na diagonal.

$$\delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \dots & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \delta_{23} & \dots & \dots & \delta_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \delta_{n1} & \delta_{n2} & \delta_{n3} & \dots & \dots & \delta_{nn} \end{pmatrix} = \begin{pmatrix} 0 & \delta_{12} & \delta_{13} & \dots & \delta_{1n} \\ & 0 & \delta_{23} & \dots & \delta_{2n} \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{pmatrix}$$

Mas nem sempre é possível associar a uma matriz quadrada δ , triangular positiva e com zeros na diagonal uma X -árvore com pesos nas arestas que a determine.

Exemplo 5. Vamos considerar uma matriz de distâncias M e verificar que não é possível representá-la por uma árvore com uma métrica associada.

$$\begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{cccc} a & b & c & d \\ \left(\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 1 & 0 & 5 & 4 \\ 2 & 5 & 0 & 3 \\ 3 & 4 & 3 & 0 \end{array} \right) \end{array}$$

A distância entre a e b é igual a 1 e a distância entre a e c é igual a 2. Como a desigualdade triangular é uma propriedade da métrica D_w , então o único caminho entre b e c teria peso ≤ 3 . Na matriz o peso do caminho entre b e c é igual a 6, logo tal condição não é satisfeita.

Vamos então assumir que existe uma árvore T tal que a distância entre os seus vértices é dada pelas entradas da matriz M , ou seja, m_{ij} representa a distância entre o vértice i e o vértice j , isto é, $M = D_w$ qualquer.

Neste exemplo a função D_w , representada na matriz, não é uma métrica, logo era fácil de verificar que não era possível representá-la numa árvore.

Vamos agora ver outro exemplo, em que a função D_w representada na matriz é uma função métrica.

Exemplo 6. Consideremos agora a matriz M representada abaixo.

$$\begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{cccc} a & b & c & d \\ \left(\begin{array}{cccc} 0 & 2 & 1 & 2 \\ & 0 & 1 & 2 \\ & & 0 & 2 \\ & & & 0 \end{array} \right) \end{array}$$

Esta matriz é positiva, simétrica e respeita a desigualdade triangular, mas, a função D_w que ela representa não pode ser associada a uma X -árvore, ou seja, não é possível atribuir pesos a uma árvore com quatro folhas, a, b, c, d , em que a distância entre os pares de folhas respeite a função D_w representada na matriz.

É fácil de verificar que uma tal função D_w não verifica a condição dos quatro pontos, pois:

$$D_w(a, b) + D_w(c, d) = 4$$

$$D_w(a, c) + D_w(b, d) = 3$$

$$D_w(a, d) + D_w(b, c) = 3$$

duas das soma são iguais, mas o seu valor é menor que a terceira soma, mas, de acordo com o Lema 3, a condição dos quatro pontos teria que ser satisfeita.

O resultado seguinte, cuja prova pode ser vista em [3], pag. 152, Teorema 7.2.6 (Tree-Metric Theorem), mostra que esta condição dos 4 pontos é necessária e suficiente para garantir que uma matriz simétrica com zeros na diagonal *tem uma representação por X -árvore*, i.e., é a matriz das distâncias D_W entre os vértices etiquetados por X de alguma certa X -árvore com uma função w de pesos nas arestas:

Teorema 2.3. *Seja δ uma função tal que $\forall x, y \in X \delta(x, x) = 0$ e $\delta(x, y) = \delta(y, x)$. Então δ tem uma representação por X -árvore se e só se satisfaz a condição dos quatro pontos.*

2.2 Enlaçar uma árvore filogenética

O *Tree Metric Theorem*, provado por independentemente de Zaretskii (1965), Simões-Pereira (1969) e Buneman (1971), afirma que numa X -árvore T , com pesos positivos nas arestas, com conjunto de folhas X e sem vértices de grau dois, fica determinado, tanto o peso de todas as arestas como a *topologia da árvore*, pelo conhecimento das distâncias entre todos os pares de folhas. E são conhecidos algoritmos, como o *Algoritmo Neighbor Joining* (ver [3, pág. 157]) que permitem reconstruir a árvore com pesos nas arestas a partir da matriz de distâncias.

A abordagem feita por Dress, Huber e Steel em [4] que vamos introduzir nesta secção, é a de estudar as propriedades que permitem garantir que certos conjuntos de pares de folhas de uma X -árvore filogenética determinem a árvore. Estes conjuntos, são denotados por \mathcal{L} e denominados por *Lassos*.

De acordo com os autores do referido artigo, uma das razões que tornam importante esta ideia é o facto de numa das possíveis aplicações em biologia, a sequenciação genética, só ser possível determinar a chamada *distância genética* entre uma parte das combinações dos *objetos biológicos* envolvidos, que no modelo matemático serão as folhas da árvore a reconstruir.

Ao longo desta secção, quando dizemos X -árvore queremos dizer X -árvore filogenética e vamos identificar o conjunto das etiquetas, X , com o conjunto das folhas da árvore.

Definição 15. Numa X -árvore, a qualquer subconjunto com dois elementos, $\binom{X}{2}$ de X chamamos *corda* e escrevemos $c = xy$. Um conjunto de cordas diz-se um *Lasso*.

Definição 16. Duas X -árvores T e T' , com duas funções peso associadas, w e w' , respectivamente, dizem-se *\mathcal{L} -isométricas* se a distância entre os pares de folhas que pertencem a \mathcal{L} forem iguais nas duas X -árvores, isto é, para todos $\{x, y\} \in \mathcal{L}$, $D_w(x, y) = D_{w'}(x, y)$ e escrevemos $(T, w) \stackrel{\mathcal{L}}{\equiv} (T', w')$.

Exemplo 7. Vamos considerar a X -árvore seguinte e o conjunto $\mathcal{L} = \{ab, bc, bd, af, ae\}$.

Consideremos a X -árvore T representada na Figura 2.19, e duas funções peso distintas w_ϵ definidas na árvore T , w_2 e w_3 , quando $\epsilon = 2$ e $\epsilon = 3$, respetivamente.

As distâncias entre todos os pares de folhas que pertencem ao Lasso \mathcal{L} são iguais nas duas árvores:

$$D_{w_2}(a, b) = 200 = D_{w_3}(a, b)$$

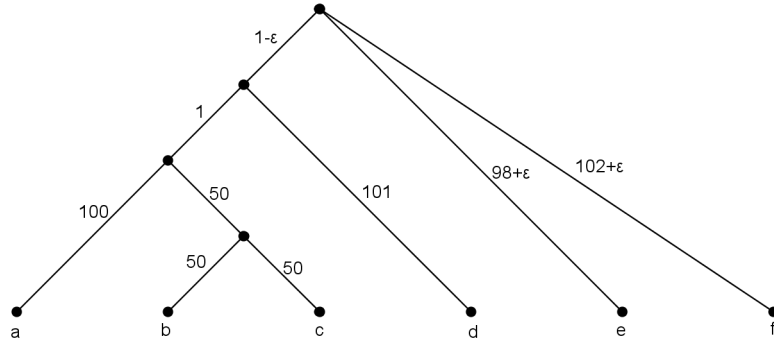


Figura 2.19: X -árvore com pesos.

$$D_{w_2}(b, c) = 100 = D_{w_3}(b, c)$$

$$D_{w_2}(b, d) = 202 = D_{w_3}(b, d)$$

$$D_{w_2}(a, f) = 204 = D_{w_3}(a, f)$$

$$D_{w_2}(a, e) = 200 = D_{w_3}(a, e)$$

logo, $(T, w_2) \stackrel{\mathcal{L}}{\equiv} (T, w_3)$, com $w_2 \neq w_3$

Exemplo 8. Vamos considerar agora duas X -árvores distintas, T e T' representadas na Figura 2.20, que não são topologicamente equivalentes, e cada uma com a sua função peso associada, w e w' , respetivamente. Consideramos também o laço $\mathcal{L} = \{ab, cd\}$.

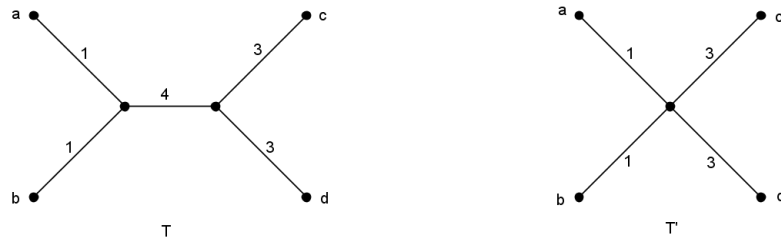


Figura 2.20: Duas X -árvores distintas.

Como no exemplo anterior, o peso do caminho entre os pares de folhas que pertencem ao laço \mathcal{L} são iguais nas duas árvores:

$$D_w(a, b) = 2 = D_{w'}(a, b)$$

$$D_w(c, d) = 6 = D_{w'}(c, d).$$

logo, $(T, w) \stackrel{\mathcal{L}}{\equiv} (T', w')$.

O peso do caminho entre os pares de folhas que não pertencem ao laço \mathcal{L} são distintos nas duas árvores e portanto para qualquer conjunto de pares \mathcal{L}' que contenha pelo menos um dos pares abaixo as duas árvores não serão \mathcal{L}' -isométricas.

$$\begin{aligned}
D_w(a, c) &= 8 \neq 4 = D_{w'}(a, c) \\
D_w(a, d) &= 8 \neq 4 = D_{w'}(a, d) \\
D_w(b, c) &= 8 \neq 4 = D_{w'}(b, c) \\
D_w(b, d) &= 8 \neq 4 = D_{w'}(b, d)
\end{aligned}$$

Definição 17. Seja T uma X -árvore e \mathcal{L} um *lasso* associado a essa mesma árvore.

Vamos denotar por $\Gamma(\mathcal{L})$ o grafo que tem como conjunto de vértices os elementos do conjunto X , e arestas os pares de folhas que pertencem ao lasso \mathcal{L} .

Por exemplo, o grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bc, bd, af, ae\}$ da Figura 2.19 é:

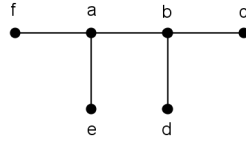


Figura 2.21: Grafo $\Gamma(\mathcal{L})$ associado a $\mathcal{L} = \{ab, bc, bd, af, ae\}$.

Definição 18. Dada uma X -árvore T e um subconjunto \mathcal{L} de $\binom{X}{2}$, definem-se três tipos de lassos:

- \mathcal{L} é *edge-weight lasso* de T se determina o peso de todas as arestas da árvore, isto é, se w e w' forem duas funções peso em T que “coincidem em \mathcal{L} ”. Isto é, tais que $(T, w) \stackrel{\mathcal{L}}{\equiv} (T, w')$ então coincidem em todas as arestas de T , ou seja, $w = w'$.
- \mathcal{L} é *lasso topológico* de T se determina a topologia da árvore, isto é, se w e w' forem duas funções peso em T e T' , respectivamente, que “coincidem em \mathcal{L} ”, isto é, tais que $(T, w) \stackrel{\mathcal{L}}{\equiv} (T', w')$, então T e T' são topologicamente equivalentes.
- \mathcal{L} *strong lasso* de T se é simultaneamente *edge-weight* e topológico; este subconjunto, \mathcal{L} , permite determinar os pesos de todas as arestas da árvore e a sua topologia.

Seja T uma X -árvore com conjunto de folhas $X = \{a, b, c, d, e, f\}$, e seja $\mathcal{L} = \{ab, ac, bc, de, df, ef\}$ um lasso associado a esta X -árvore. É fácil verificar que \mathcal{L} não é um *edge-weight lasso*.

Como já foi visto na Definição 18, \mathcal{L} é um *edge-weight lasso* se w e w' forem duas funções peso em T tal que $(T, w) \stackrel{\mathcal{L}}{\equiv} (T, w')$, então coincidem em todas as arestas de T , ou seja, $w = w'$.

É conhecido o comprimento dos caminhos que ligam quaisquer duas folhas do conjunto: $\{ab\}, \{ac\}, \{bc\}, \{de\}, \{df\}, \{ef\}$. No entanto, não é possível determinar a distância entre uma folha do conjunto $\{abc\}$ e do conjunto $\{def\}$. Consideremos e_k uma aresta do caminho que ligue duas folhas dos conjuntos $\{abc\}$ e $\{def\}$ e que não pertença aos caminhos que ligam os pares de folhas do conjunto \mathcal{L} . Essa aresta pode tomar

qualquer valor, o que não garante que $w = w'$, logo $\mathcal{L} = \{ab, ac, bc, de, df, ef\}$ não é um *edge-weight* lasso para a X -árvore T .

Por fim, vamos considerar a X -árvore representada na Figura 2.22 com conjunto de folhas $X = \{a, b, c, d, e, f\}$, e seja $\mathcal{L} = \{ab, ac, bc, de, df, ef, ae, bd, cf\}$ um lasso associado a T .

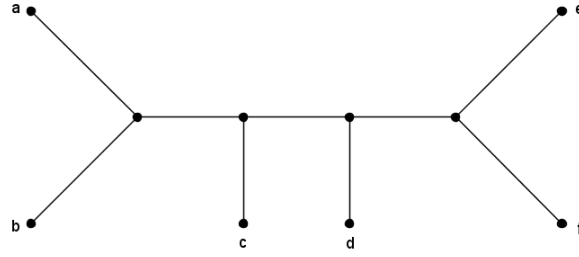


Figura 2.22: X -árvore T .

Vamos verificar que \mathcal{L} é um *edge-weight* lasso para a X -árvore T , uma vez que é possível determinar, para qualquer função peso w de T , os valores $D(x, y) := D_w(x, y)$, para todas as cordas $xy \in \binom{X}{2}$.

As distâncias entre quaisquer $xy \in \mathcal{L}$ são conhecidas. Vamos verificar se, para todas as cordas que não pertençam ao lasso \mathcal{L} , é possível determinar o valor de $D(x, y)$. Por exemplo:

$$D(a, d) = D(a, c) + D(b, d) - D(b, c)$$

depois de conhecida a distância entre as folhas a e d , podemos determinar:

$$D(b, e) = D(a, e) + D(b, d) - D(a, d).$$

Analogamente, determinamos a distância entre os pares de folhas restantes:

$$D(c, e) = D(d, e) + D(c, f) - D(d, f)$$

$$D(a, f) = D(c, f) + D(a, e) - D(c, e)$$

$$D(b, f) = D(b, e) + D(a, f) - D(a, e)$$

$$D(c, d) = D(c, e) + D(a, d) - D(a, e).$$

Para uma certa função peso w , é possível determinar a distância entre todos os pares de folhas, então $\mathcal{L} = \{ab, ac, bc, de, df, ef, ae, bd, cf\}$ é um *edge-weight* lasso.

No entanto, \mathcal{L} não é um lasso topológico, pois existem, pelo menos, duas X -árvores que não são topologicamente equivalentes com o mesmo peso entre todos os pares de folhas da X -árvore T . Na Figura 2.23 estão representadas duas X -árvores, T_1 e T_2 , com a mesma distância entre todos os pares de folhas de \mathcal{L} , mas não são topologicamente equivalentes.

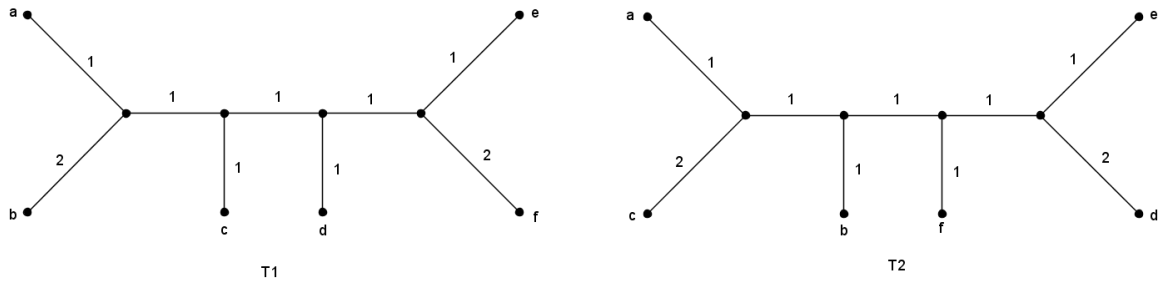


Figura 2.23: Duas árvores que não são topologicamente equivalentes: T_1 e T_2 .

Admitamos que existe uma função $\phi : V(T_1) \rightarrow V(T_2)$, que é um isomorfismo nas duas X -árvores.

Para as duas X -árvores, T_1 e T_2 , considere-se:

- x e y dois vértices interiores de T_1 ,
- ax a aresta com extremidades a e x ; cy a aresta com extremidade c e y ,
- $\phi(x)$ e $\phi(y)$ as imagens de x e y em T_2 , respetivamente,

representado na Figura 2.24.

Assim sendo, a aresta ax tem que ter por imagem uma aresta que tem extremidade em a , e outra extremidade em $\phi(x)$ (imagem de x em T_2).

Do mesmo modo, a aresta cy tem que ter por imagem uma aresta que tem extremidade em c , e outra extremidade em $\phi(y)$ (imagem de y em T_2).

Com esta construção, na X -árvore T_2 , temos que $\phi(x) = \phi(y)$, o que contradiz o item 1 da Definição 13, logo $T_1 \not\cong T_2$.

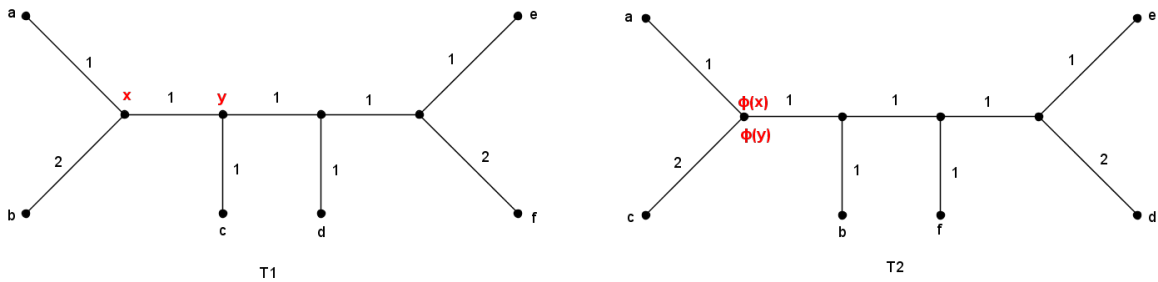


Figura 2.24: T_1 e T_2 .

Teorema 2.4. Se $n \geq 4$ e \mathcal{L} é lasso topológico de T , então $\Gamma(\mathcal{L})$ deve ser conexo.

Demonstração. Suponha que existe uma bipartição de X em dois subconjuntos disjuntos não vazios, A e B , tal que \mathcal{L} não contém nenhuma corda formada por ab onde $a \in A$ e $b \in B$. Considera-se w qualquer função de peso em T . Consideremos $T|_A$ e $T|_B$ as duas árvores obtidas através da restrição de T a A e B , respetivamente, e consideremos $w|_A$ e $w|_B$, as restrições de w a estas subárvores, as suas funções de peso.

Podemos então construir uma X -árvore T' com função de peso w' tal que T' não é topologicamente equivalente a T apesar de se verificar $T|_A \cong T'|_A$ e $T|_B \cong T'|_B$, bem como $w_A = w'_A$ e $w_B = w'_B$ e, por conseguinte, (T, w) e (T', w') serem \mathcal{L} -isométricas.

Isso pode ser feito pela “fusão” de $T|_A$ e $T|_B$ por arestas convenientemente escolhidas.

□

A Figura 2.25 representa a X -árvore com conjunto $X = \{a, b, c, d, e, f\}$, conjunto $\mathcal{L} = \{ab, bc, ac, de, ef, fd\}$ e o grafo $\Gamma(\mathcal{L})$. Como o grafo é desconexo, então o lasso \mathcal{L} não é topológico.

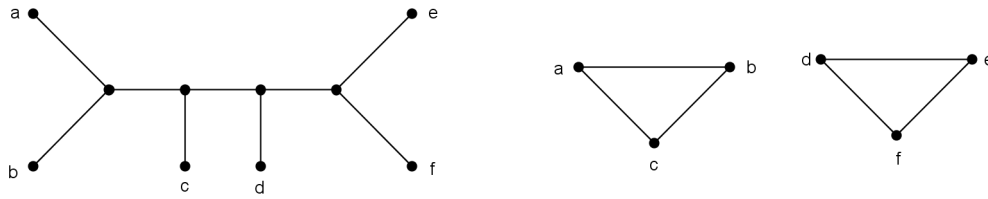


Figura 2.25: X -árvore com conjunto $X = \{a, b, c, d, e, f\}$ e o seu grafo $\Gamma(\mathcal{L})$ associado ao conjunto $\mathcal{L} = \{ab, bc, ac, de, ef, fd\}$.

Teorema 2.5. Se \mathcal{L} é *weight-lasso* de T , então $\Gamma(\mathcal{L})$ é fortemente não bipartido, i.e., todas as componentes conexas do grafo $\Gamma(\mathcal{L})$ são não bipartidas.

Demonstração. Suponha que $\Gamma(\mathcal{L})$ contém uma componente conexa que é bipartida: seja $Y \subseteq X$ o conjunto de vértices dessa componente e sejam Y^+ e Y^- , os dois conjuntos de uma bipartição de Y tais que não existem arestas do grafo a ligar vértices Y^+ entre si, nem de Y^- entre si.

Dada qualquer função de peso de T , podemos adicionar uma constante suficientemente pequena, τ , aos pesos de todas as arestas que contém uma folha em Y^+ e subtrair a mesma quantidade a todos os pesos das arestas que contém uma folha em Y^- , sem alterar a distância entre quaisquer duas folhas $x, x' \in X$ com $xx' \in \mathcal{L}$. Sendo assim, o lasso não é *edge-weight* lasso. □

Na Figura 2.26 está representada uma X -árvore com conjunto $X = \{a, b, c, d, e\}$. Consideremos $\mathcal{L} = \{ab, ac, bc, de\}$ e o seu grafo $\Gamma(\mathcal{L})$ que como verificámos na figura é um grafo desconexo.

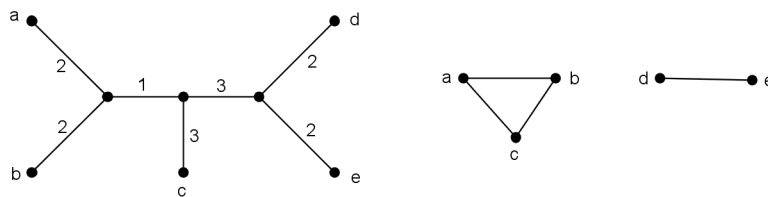


Figura 2.26: X -árvore com conjunto $X = \{a, b, c, d, e\}$ e o grafo $\Gamma(\mathcal{L})$ associado ao conjunto $\mathcal{L} = \{ab, ac, bc, de\}$.

Consideremos a componente conexa bipartida do grafo, $\{de\}$. Temos então os conjuntos de bipartição $Y^+ = \{d\}$ e $Y^- = \{e\}$ e somamos τ a todas as arestas incidentes nas folhas que pertencem ao conjunto Y^+ e subtraímos τ a todas as arestas que incidem nas folhas que pertencem ao conjunto Y^- , ilustrado na Figura 2.27.

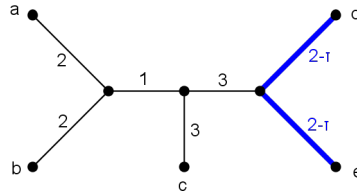


Figura 2.27: X-árvore com conjunto $X = \{a, b, c, d, e\}$ com diferentes pesos nas arestas.

Como verificamos na Figura 2.27, a distância entre todos os pares de folhas que pertencem ao laço \mathcal{L} mantêm-se nas duas árvores, variando o valor τ obtemos várias X-árvores distintas e a distância entre os pares de folhas do laço \mathcal{L} não se altera. Concluimos então que o laço \mathcal{L} não é *edge-weight* laço.

Corolário 1. Em particular, $\Gamma(\mathcal{L})$ é conexo e não bipartido se \mathcal{L} é *strong*-laço para T .

Demonstração. Pelo Teorema 2.4 sabe-se que se $\Gamma(\mathcal{L})$ é um grafo conexo, então \mathcal{L} é um laço topológico. Por outro lado, pelo Teorema 2.5 sabe-se que se $\Gamma(\mathcal{L})$ é fortemente não bipartido, então \mathcal{L} é um *edge-weight* laço.

Assim, se $\Gamma(\mathcal{L})$ é um grafo conexo e fortemente não bipartido, então o laço \mathcal{L} é um *edge-weight* laço e laço topológico, ou seja, um *strong* laço. \square

Capítulo 3

Matroides

Neste capítulo introduzimos uma estrutura combinatória que será associada no capítulo seguinte às X -árvores filogenéticas, um *matroide*.

O matroide pode ser introduzido de diversas formas que podem estar mais diretamente relacionadas com alguns dos objetos matemáticos a que esta estrutura pode ser associada: por exemplo, o conjunto de *independentes*, o das *bases* ou o dos *circuitos*. O conceito de base de um espaço vetorial e o de árvore geradora de um grafo, são dois dos mais conhecidos exemplos diretamente ligados a esta estrutura combinatória.

Ela é definida num conjunto finito e não vazio, dito o seu *conjunto de suporte*, que no caso em estudo no próximo capítulo será $\binom{X}{2}$, visto como o conjunto dos pares de folhas de uma X -árvore. Neste caso, será usada uma função, denominada por *função rank*, para introduzir o matroide associado a uma qualquer X -árvore.

As provas de alguns dos resultados aqui apresentados bem como outras propriedades e resultados gerais em Teoria de Matroides podem ser consultados, por exemplo, em [1], [2] e [6].

3.1 Definição e Propriedades

Definição 19. Seja M um conjunto finito e \mathcal{I} um conjunto de partes de M , que satisfaçam as seguintes condições:

- (i) $\emptyset \in \mathcal{I}$;
- (ii) Se $X \in \mathcal{I}$ e $Y \subseteq X$ então $Y \in \mathcal{I}$;
- (iii) Se I_1 e $I_2 \in \mathcal{I}$ e $|I_1| < |I_2|$ então existe um elemento $b \in I_2 \setminus I_1$ tal que $I_1 \cup \{b\} \in \mathcal{I}$.

Dizemos então que $\mathcal{M} = (M, \mathcal{I})$ é um *matroide*.

Se $\mathcal{M} = (M, \mathcal{I})$ é um matroide, dizemos que \mathcal{M} é um matroide sobre M ou que M é o *conjunto de suporte* de \mathcal{M} .

Os elementos de \mathcal{I} , também denotado por $\mathcal{I}(\mathcal{M})$, dizem-se os *independentes* de \mathcal{M} .

Os subconjuntos de M que não estão em \mathcal{I} são ditos os *dependentes* de \mathcal{M} .

Observe-se o que resulta da Definição 19:

- o conjunto de independentes é não vazio porque $\emptyset \in \mathcal{I}$.
- Como $\mathcal{I} \neq \emptyset$ e M é finito existe sempre pelo menos um conjunto independente maximal para a inclusão:

Vamos supor x_0 um elemento do conjunto M e I_0 um subconjunto independente.

Se existir um elemento $x_1 \in M \setminus \{x_0\}$ tal que $I_1 = I_0 \cup \{x_1\}$ é um subconjunto independente, isto implica que existe um subconjunto independente que contém estritamente o anterior; caso contrário, I_0 é um conjunto independente maximal. Em geral, se I_{i-1} é um conjunto independente, ou podemos acrescentar um elemento $x_i \in M \setminus I_{i-1}$ tal que $I_i = I_{i-1} \cup \{x_i\}$ é um conjunto independente ou $I_i = I_{i-1} \cup \{x_i\}$ é dependente, para todo $x_i \in M \setminus I_{i-1}$, e nesse caso I_{i-1} é independente maximal. Como M é um conjunto finito e existe pelo menos um I_0 isto implica que existe um k tal que I_k é um subconjunto independente maximal.

- Os subconjuntos independentes maximais são ditos *bases* do matroide.

Lema 4. Quaisquer dois conjuntos independentes maximais têm o mesmo número de elementos.

Demonstração. Sejam I_1 e I_2 dois conjuntos independentes maximais de um matroide \mathcal{M} e admitamos por redução ao absurdo que $|I_1| \neq |I_2|$; sem perda de generalidade podemos admitir que $|I_1| < |I_2|$. Pela Definição de matroide, existe um elemento $b \in I_2 \setminus I_1$ tal que $I_1 \cup \{b\}$ é independente. Sabemos que $I_1 \cup \{b\} \supsetneq I_1$ porque $b \in I_2 \setminus I_1$, o que contraria o facto de I_1 ser maximal.

Logo, nem $|I_1| < |I_2|$ nem $|I_2| < |I_1|$. Assim sendo $|I_1| = |I_2|$.

□

Decorre do lema anterior que, quaisquer duas bases de um matroide \mathcal{M} , têm o mesmo número de elementos.

Teorema 3.1. Seja \mathcal{B} uma coleção de subconjuntos de M . Se \mathcal{B} é a coleção de bases de um matroide em M , então, satisfaz as seguintes condições:

- $\mathcal{B} \neq \emptyset$;
- Se A e B são membros distintos de \mathcal{B} e $a \in A \setminus B$, então existe um elemento $b \in B \setminus A$ tal que $A \setminus \{a\} \cup \{b\} \in \mathcal{B}$.

Demonstração. Como já foi visto anteriormente, a coleção das bases de um matroide é não vazia.

Sejam A e $B \in \mathcal{B}$ duas bases quaisquer.

As bases de um matroide são os subconjuntos independentes maximais, então, pelo Lema 4 sabemos que $|A| = |B|$.

Consideremos $x \in A \setminus B$. Como $A \setminus \{x\} \subseteq A$ e $A \in \mathcal{I}$, então pela Definição 19, $A \setminus \{x\} \in \mathcal{I}$.

Dado que $|A \setminus \{x\}| = |A| - 1 = |B| - 1 < |B|$, pela Definição 19 sabemos que existe $y \in B \setminus (A \setminus \{x\})$ tal que $(A \setminus \{x\}) \cup \{y\} \in \mathcal{I}$.

Falta agora provar que este conjunto, além de independente, é maximal.

Como $|(A \setminus \{x\}) \cup \{y\}| = |A| = |B|$ e $\{x\} \neq \{y\}$, então $(A \setminus \{x\}) \cup \{y\}$ é maximal, se não fosse teríamos independentes maximais com cardinais distintos, o que contraria o Lema 4. \square

O recíproco do Teorema é verdade, ou seja, se uma coleção \mathcal{B} de subconjuntos de um conjunto M satisfaz as duas propriedades, então o par $\mathcal{M} = (M, \mathcal{I})$ é um matroide, onde $\mathcal{I} = \{I \in \mathcal{P}(M) : I \subseteq B \text{ para algum } B \in \mathcal{B}\}$ e $\mathcal{P}(M)$ é o conjunto das partes de M .

A prova do recíproco do Teorema pode ser vista na página 8 de [2].

De seguida, são apresentados alguns exemplos de matroides.

Exemplo 9 (Matroide Livre). Seja M um conjunto com n elementos tal que $M \in \mathcal{I}$. Neste caso, todos os subconjuntos de E são independentes, ou seja, $\mathcal{I} = \mathcal{P}(M)$. O par $\mathcal{M} = (M, \mathcal{P}(M))$ define um matroide dito *matroide livre*.

Exemplo 10. Sejam n e k dois inteiros não negativos tais que $k \leq n$. Seja E o conjunto com n elementos e \mathcal{B} a coleção de subconjuntos de M com k elementos.

A família \mathcal{B} verifica os axiomas das bases de um matroide e assim o par $\mathcal{M} = (M, \mathcal{B})$ define um matroide chamado *matroide uniforme*, denotado por $U_{n,k}$.

A qualquer grafo podemos associar um matroide.

Definição 20 (*Matroide Gráfico*). Consideremos um grafo conexo G e seja $E(G)$ o conjunto de arestas de G .

A este grafo podemos associar o matroide, $M_G = (E_G, \mathcal{I}_G)$, com conjunto de suporte igual ao conjunto das arestas do grafo G , E_G , e conjunto de independentes os conjuntos acíclicos de arestas de G , i.e., os conjuntos de arestas dos seus subgrafos acíclicos, $\mathcal{I}_G = \{I \subseteq E_G : G[I] \text{ é acíclico}\}$.

Seja \mathcal{B}_G o conjunto formado pelos conjuntos de arestas acíclicos maximais; provamos em [6] que os elementos deste conjunto são as árvores geradoras do grafo G . Como referimos anteriormente, as bases dos matroides tem propriedades básicas que podem ser tomadas como axiomas para um matroide. Deste modo, caracterizamos o matroide através do seu conjunto de bases.

Teorema 3.2. *Se G é um grafo conexo, então $M_G = (E_G, \mathcal{B}_G)$ é um matroide.*

Demonstração. Consideremos A e B duas árvores geradoras distintas do grafo G , A e $B \in \mathcal{B}_G$ e seja $a \in A \setminus B$ e $b \in B \setminus A$.

Queremos provar que $A \cup \{b\} \setminus \{a\} \in \mathcal{B}_G$, ou seja, que é uma árvore geradora de G .

Consideremos uma aresta $b \in B \setminus A$, $A \cup \{b\}$ contém um único ciclo, C_b , ver secção 2.1.2. Como B é uma árvore, então $C_b \not\subseteq B$.

Sabemos então, que existe uma aresta $a \in A \setminus B$ tal que $A \cup \{b\} \setminus \{a\}$ é acíclico, ou seja, uma árvore.

Como $|A \cup \{b\} \setminus \{a\}| = |A| = |B|$ então $A \cup \{b\} \setminus \{a\}$ é uma árvore geradora, pois todas as árvores geradoras tem a mesma cardinalidade.

Assim, M_G é um matroide. \square

Exemplo 11. Na Figura 3.1 está representado o grafo $G = K_4$, sendo K_4 o grafo completo com quatro vértices.

Como acima já foi definido, o conjunto de suporte é o conjunto das arestas do grafo, ou seja, $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$. O conjunto dos independentes do matroide são as florestas do grafo.

Assim, definimos o matroide $M_G = (E_G, \mathcal{I}_G)$.

O conjunto das bases deste matroide são as florestas maximais do grafo G . Como G é um grafo conexo, as bases do matroide são as árvores geradoras:

$$\mathcal{B}_G = \{\{e_1, e_2, e_5\}, \{e_1, e_2, e_6\}, \{e_1, e_2, e_4\}, \{e_1, e_3, e_5\}, \{e_1, e_3, e_4\}, \\ \{e_1, e_3, e_6\}, \{e_4, e_3, e_6\}, \{e_3, e_2, e_6\}, \{e_3, e_2, e_5\}, \{e_3, e_2, e_4\}, \dots\}$$

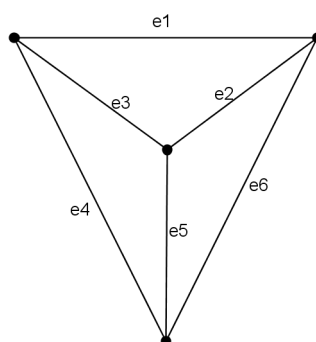


Figura 3.1: Grafo completo com quatro vértices, K_4 .

Definição 21. Um *circuito* de um matroide é um subconjunto C de M , dependente minimal, isto é, tal que C é dependente e $\forall x \in C$, $C \setminus \{x\}$ é independente. A coleção dos circuitos do matroide é denotada por \mathcal{C} .

Como já referimos, a coleção das bases tem propriedades simples que podem ser tomadas como axiomas para um matroide. O mesmo acontece com os circuitos.

Teorema 3.3. *Seja M um conjunto finito; $\mathcal{C} \subseteq \mathcal{P}(M)$ é a coleção dos circuitos de um matroide \mathcal{M} , denotada por \mathcal{C} , se e só se tem as seguintes propriedades:*

- (i) $\emptyset \notin \mathcal{C}$;
- (ii) Se $C_1, C_2 \in \mathcal{C}$, com $C_1 \subseteq C_2$ então $C_1 = C_2$;
- (iii) Se $C_1, C_2 \in \mathcal{C}$, com $C_1 \neq C_2$ e $a \in C_1 \cap C_2$ então existe $C_3 \in \mathcal{C}$ tal que $C_3 \subseteq (C_1 \cup C_2) \setminus a$.

A prova do Teorema 3.3 pode ser vista na página 23 [1].

De modo a ilustrar os circuitos de um matroide associado a um grafo, vamos considerar o grafo G representado na Figura 3.2. Seja $\mathcal{M} = M_G$, o matroide gráfico associado ao grafo G . O par $\mathcal{M} = (E, \mathcal{C})$ que representa o matroide, tem como conjunto de suporte E as arestas e \mathcal{C} os ciclos do grafo G . Ou seja, $E = \{e_1, e_2, e_3, e_4, e_5\}$ e $\mathcal{C} = \{\{e_3\}, \{e_1, e_4\}, \{e_1, e_2, e_5\}, \{e_2, e_4, e_5\}\}$.

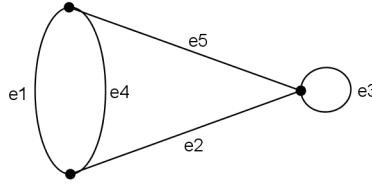


Figura 3.2: Grafo G .

3.2 Circuitos e rank num matroide

Definição 22 (Matroide associado a uma matriz). Dada uma matriz $A \in M_{m \times n}(\mathbb{R})$, consideremos que cada coluna da matriz é um vetor de \mathbb{R}^m . As colunas desta matriz são denotadas por u_i , com $i = 1, 2, \dots, n$.

Seja $\mathcal{M} = (M, \mathcal{I})$ o matroide onde $M = \{1, 2, 3, \dots, n\}$ é o conjunto dos índices da colunas da matriz A . Os independentes de \mathcal{M} , elementos do conjunto \mathcal{I} , são definidos por:

$$\{i_1, \dots, i_k\} \in \mathcal{I} \Leftrightarrow \{u_{i_1}, \dots, u_{i_k}\} \text{ é linearmente independente,}$$

ou seja, os índices dos conjuntos de colunas da matriz, que são linearmente independentes, formam os subconjuntos independentes do matroide.

Este matroide é denominado por *matroide vetorial ou linear*.

Teorema 3.4. Sendo M e \mathcal{I} os subconjuntos definidos anteriormente, então $\mathcal{M} = (M, \mathcal{I})$ define um matroide.

A prova do Teorema 3.4 pode ser vista nas páginas 7 e 8 de [6].

Exemplo 12. Seja A a matriz $A = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$ sobre \mathbb{R} . Os elementos do conjunto de suporte $M = \{1, 2, 3, 4, 5\}$, são os índices dos conjuntos das colunas da matriz A . O conjunto dos independentes do matroide é:

$$\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{4\}, \{5\}, \{1, 2\}, \{1, 5\}, \{2, 4\}, \{2, 5\}, \{4, 5\}\}.$$

Assim, o conjunto de dependentes do matroide é:

$$\{\{3\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{3, 5\}\} \cup \{X \subseteq M : |X| \geq 3\}.$$

As bases do matroide são os conjuntos independentes maximais, neste caso: $\mathcal{B} = \{\{1, 2\}, \{1, 5\}, \{2, 4\}, \{2, 5\}, \{4, 5\}\}$.

Exemplo 13. Seja A a matriz $A = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$.

O conjunto de suporte $M = \{1, 2, 3, 4\}$, é o conjunto dos índices das colunas da matriz, e \mathcal{I} é o subconjunto de partes de M , ou seja, o conjunto dos índices dos conjuntos de colunas da matriz que são linearmente independentes.

$$\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \\ \{2, 4\}, \{1, 2, 3\}, \{1, 3, 4\}, \{2, 3, 4\}\}$$

As bases deste matroide são $\mathcal{B} = \{\{1, 2, 3\}, \{1, 3, 4\}, \{2, 3, 4\}\}$, ou seja, o conjunto dos independentes maximais.

Seja \mathcal{M} um matroide em M e A um subconjunto de M . A função *rank* é definida no conjunto das partes de M do seguinte modo:

$$\begin{array}{ccc} \text{rank}: & \mathcal{P}(M) & \longrightarrow \\ & X & \mapsto \text{rank}(X) = \{\max_{I \in \mathcal{I}, I \subseteq X} |I| : I \in \mathcal{I}, I \subseteq X\} \end{array} \cdot$$

Esta função verifica as seguintes propriedades:

- $\text{rank}(A) \in \mathbb{Z}_0^+$;
- $\text{rank}(A) \leq |A|$;
- Para quaisquer dois subconjuntos A e B de M :
 $\text{rank}(A \cup B) + \text{rank}(A \cap B) \leq \text{rank}(A) + \text{rank}(B)$;
- Para qualquer subconjunto A e elemento x :
 $\text{rank}(A) \leq \text{rank}(A \cup \{x\}) \leq \text{rank}(A) + 1$.

Como já referimos, a coleção das bases ou dos circuitos, têm propriedades simples que podem ser tomadas como axiomas para um matroide, o mesmo acontece com a função *rank*.

Com as propriedades acima é fácil verificar que a função *rank* é monótona, ou seja, se $A \subset B$ então $\text{rank}(A) \leq \text{rank}(B)$.

Sejam A e B dois subconjuntos de E .

Usando as propriedades acima sabemos que: $\text{rank}(A \cup \{x\}) \leq \text{rank}(A) + 1$.

Se $A = B$ então $\text{rank}(A) = \text{rank}(B)$. Se $A \subset B$ então existe um conjunto de elementos $\{x_1, x_2, \dots, x_k\} \in B \setminus A$, assim sendo:

$$\text{rank}(A \cup \{x_1, x_2, \dots, x_k\}) \leq \text{rank}(A) + |\{x_1, x_2, \dots, x_k\}| \leq \text{rank}(B).$$

Assim, usando a função *rank*, definimos as seguintes coleções:

- $\mathcal{I} = \{I \subseteq M : \text{rank}(I) = |I|\}$ é a coleção dos conjuntos independentes do matroide;
- $\mathcal{G} = \{L \subseteq M : \text{rank}(L) = \text{rank}(M)\}$ é a coleção dos conjuntos de *geradores*;
- \mathcal{B} é a coleção das bases do matroide, ou seja, os conjuntos maximais de \mathcal{I} ;
- \mathcal{C} é a coleção dos circuitos, ou seja, os conjuntos minimais de $\mathcal{P}(M) - \mathcal{I}$.

Exemplo 14. Considerando o grafo da Figura 3.2, concluímos que: $rank(\{e_3\}) = 0$, $rank(\{e_1, e_4\}) = 1$, $rank(\{e_1, e_2, e_5\}) = 2$, $rank(\{e_1, e_4, e_5\}) = 2$ e $rank(\{e_1, e_2, e_5\}) = 2$.

Exemplo 15. Considerando agora o matroide do Exemplo 12. Temos os cinco vetores que são as colunas da matriz A :

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} u_4 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_5 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Sabemos então que: $rank(\{u_1\}) = 1$, $rank(\{u_3\}) = 0$, $rank(\{u_1, u_2\}) = 2$, $rank(\{u_1, u_3\}) = 1$, $rank(\{u_1, u_2, u_3\}) = 2$, $rank(\{u_1, u_2, u_3, u_4\}) = 2$ e $rank(\{u_1, u_2, u_3, u_5\}) = 2$.

Definição 23. Seja V um espaço vetorial, $X = \{x_1, x_2, \dots, x_n\}$ um conjunto de geradores de V e $Y \in \mathcal{P}(X)$ um conjunto de vetores contido em X . Consideremos uma função f , definida do seguinte modo:

$$f: \begin{array}{ccc} \mathcal{P}(X) & \longrightarrow & \mathbb{Z}_0^+ \\ Y & \mapsto & \dim(< Y >) \end{array},$$

onde $< Y >$ representa o subespaço vetorial gerado pelos vetores de $Y = \{y_1, y_2, \dots, y_n\}$, ou seja, o conjunto de todas as combinações lineares possíveis dos y_i .

Queremos provar que esta função f , verifica as propriedades de uma função rank (definida anteriormente). Consideremos V um espaço vetorial, $X = \{x_1, x_2, \dots, x_n\}$ um conjunto de geradores e Y um conjunto de vetores contido em X .

Se $B = \{b_1, b_2, \dots, b_k\}$ é uma base do espaço vetorial V , dizemos que V tem dimensão k , e escrevemos $\dim(V) = k$.

Um subconjunto de vetores $S = \{s_1, \dots, s_k\}$, de um subespaço vetorial V diz-se *linearmente dependente*, se existe um subconjunto finito de escalares, λ_i , com $i = \{1, \dots, k\}$, em que pelo menos um não é nulos, tais que $\sum_{i=1}^k \lambda_i y_i = 0$. O conjunto S diz-se *linearmente independente* se para qualquer subconjunto finito de escalares λ_i se tem $\sum_{i=1}^k \lambda_i = 0 \Rightarrow \lambda_i = 0 \forall i \in \{1, \dots, k\}$.

Vamos verificar se f satisfaz as propriedades:

1. $\dim(< Y >) \in \mathbb{Z}_0^+$, porque é uma dimensão de um subespaço vetorial;
2. A dimensão de um subespaço vetorial Y , é igual ao cardinal máximo de um conjunto de vetores linearmente independentes. A $\dim(< Y >)$ é no máximo igual a $|Y|$, e caso os vetores $Y = \{y_1, y_2, \dots, y_k\}$ sejam linearmente dependentes, a $\dim(< Y >)$ será igual à cardinalidade do maior subconjunto de vetores linearmente independentes. Logo, $\dim(< Y >) \leq |Y|$;
3. Sejam $A, B \subseteq X$.

Sabemos que

$$\dim(< A > + < B >) = \dim < A > + \dim < B > - \dim(< A > \cap < B >).$$

Como $< A \cup B > = < A > + < B >$, temos que:

$$f(A \cup B) = f(A) + f(B) - \dim(< A > \cap < B >).$$

Como $< A \cap B > \subseteq < A > \cap < B >$ temos também que:

$$f(A \cap B) = \dim < A \cap B > \leq \dim(< A > \cap < B >),$$

logo

$$f(A \cup B) + f(A \cap B) = f(A) + f(B) - \dim(< A > \cap < B >) + f(A \cap B) \leq f(A) + f(B).$$

4. Por outro lado, $f(A \cup \{x\}) = \dim < A \cup \{x\} > \geq \dim < A > = f(A)$,

$$\text{logo } f(A) \leq f(A \cup \{x\}).$$

Fixando $B = \{x\}$ em 3 temos também:

$$f(A \cup \{x\}) \leq f(A) + f(\{x\}) - f(A \cap \{x\}) \leq f(A) + 1,$$

já que $f(\{x\}) = \dim < \{x\} > \leq 1$ e $f(A \cap \{x\}) \geq 0$.

Capítulo 4

Matroide associado aos lassos de uma árvore filogenética

Neste capítulo vamos estudar um matroide associado a uma árvore filogenética. Dada uma X -árvore T , sem vértices de grau dois, vamos associar-lhe um matroide $\mathcal{M}(T)$ com conjunto de suporte $\binom{X}{2}$, onde X é o conjunto das folhas de T .

Baseando-nos por [4], vamos descrever este matroide e investigar as suas propriedades. A motivação deste estudo é a sua relevância para o problema da reconstrução de árvores filogenéticas.

No capítulo 2 já vimos que, conhecendo algumas das distâncias (relativamente a uma função peso) entre determinados pares de folhas de uma árvore, pode ser possível determinar a topologia da árvore, os pesos das suas arestas, ou ambos.

Vamos assim dirigir a nossa atenção para uma X -árvore T fixa (sem pesos), e um subconjunto de cardinalidade mínima \mathcal{L} de $\binom{X}{2}$, para o qual as distâncias entre todos os pares de folhas $x, y \in X$, com $xy \in \mathcal{L}$, em relação a alguma função de peso w de T , é suficiente para determinar todas as outras distâncias relativas a w , ou seja, um *edge-weight lasso*.

Ao longo deste capítulo vamos assumir que todas as X -árvores $T = (V, E)$, com conjunto de vértices V , conjunto de folhas $X \subseteq V$, e conjunto de arestas $E \subseteq \binom{V}{2}$, não têm vértices de grau dois.

4.1 Definição do matroide

O matroide que vamos definir neste capítulo tem como conjunto de suporte o conjunto dos pares de folhas de uma X -árvore T e vai ser definido a partir da sua função rank.

Lembramos que, a cada par de folhas $xy \in \binom{X}{2}$, associamos uma função $\lambda_{xy} : \mathbb{R}^E \rightarrow \mathbb{R}$ definida por $\lambda_{xy}(w) = \sum_{e \in E} w(e) \delta_{e|xy}$, onde $\begin{cases} \delta_{e|xy} := 1 \text{ se } e \in E(x|y) \\ \delta_{e|xy} := 0 \text{ caso contrário} \end{cases}$, com $E(x|y)$ o conjunto das arestas que pertencem ao caminho que liga x a y .

Consideremos uma função que vai ser definida nas partes de $\binom{X}{2}$. Seja $\mathcal{L} \subseteq \binom{X}{2}$, então definimos a função f do seguinte modo:

$$f(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = \dim(\langle \lambda_{xy} : xy \in \mathcal{L} \rangle),$$

onde $\langle \lambda_{xy} : xy \in \mathcal{L} \rangle$ designa o subespaço de $\widehat{\mathbb{R}^E}$ (o dual de \mathbb{R}^E) gerado pelas funções λ_{xy} , com $xy \in \mathcal{L}$.

Teorema 4.1. *Seja $\binom{X}{2}$ o conjunto dos pares de folhas de uma X -árvore e f a função definida acima. Esta função verifica as propriedades da função rank de um matroide.*

A demonstração deste Teorema decorre da observação, feita no capítulo 3, onde demonstrámos que a função f verifica as propriedades da função rank.

Decorre do Teorema 4.1 que a função f determina univocamente um matroide com conjunto de suporte $\binom{X}{2}$ e função rank igual a f . Designaremos esse matroide por $\mathcal{M}(T)$.

Dado $e \in E$, seja w_e a função definida no conjunto das arestas da árvore T do seguinte modo:

$$w_e: E \longrightarrow \mathbb{R} \\ f \mapsto \delta_{e,f} = \begin{cases} 1 & \text{se } e = f \\ 0 & \text{se } e \neq f \end{cases}.$$

Por exemplo, w_{e_1} é uma função que vale 1 na aresta e_1 e 0 nas restantes; w_{e_2} é uma função que vale 1 na aresta e_2 e 0 nas restantes; e assim sucessivamente.

Observe-se que o conjunto das funções $\{w_e : e \in E\}$ é uma base de \mathbb{R}^E . Tendo em conta esta base, podemos associar a cada função $\lambda_{xy} : \mathbb{R}^E \longrightarrow \mathbb{R}$ um vetor de $\mathbb{R}^{|E|}$, nomeadamente a matriz de λ_{xy} relativamente à base $\{w_e : e \in E\}$ de \mathbb{R}^E e à base canónica de \mathbb{R} :

$$\lambda_{xy} \mapsto (\lambda_{xy}(w_{e_1}) \lambda_{xy}(w_{e_2}) \dots \lambda_{xy}(w_{e_k})), \text{ com } k = |E| \text{ e } E = \{e_1, e_2, \dots, e_k\}.$$

Observe-se também que este vetor determina λ_{xy} . A cada lasso \mathcal{L} podemos associar a matriz $M_{\mathcal{L}}$, cujas linhas são os vetores associados a λ_{xy} , para cada $xy \in \mathcal{L}$.

Observe-se que o rank de \mathcal{L} é igual à característica da matriz $M_{\mathcal{L}}$.

Na Figura 4.1, estão representados três tipos específicos de X -árvores que irão desempenhar um papel importante neste trabalho.

Definição 24. (i) As *árvores estrela*, são X -árvores que têm apenas um vértice interior e, por conseguinte, são equivalentes à árvore $T^*(X) := (V^*(X), E^*(X))$, com conjunto de folhas X , vértices $V^*(X) := X \cup \{*\}$ e conjunto de arestas $E^*(X) := \{*x : x \in X\}$, onde “ $*$ ” denota o único vértice interior de $T^*(X)$.

(ii) As *árvores quarteto*, são X -árvores binárias que têm quatro folhas. $T_{ab|cd}$ representa o *quarteto* com conjunto de folhas $\{a, b, c, d\}$, cuja aresta central é denotada por $e_{ab|cd}$, e separa as folhas a, b das folhas c, d .

(iii) As *árvores lagarta*, são X -árvores binárias que contêm um caminho e todas as suas folhas estão à mesma distância desse caminho.

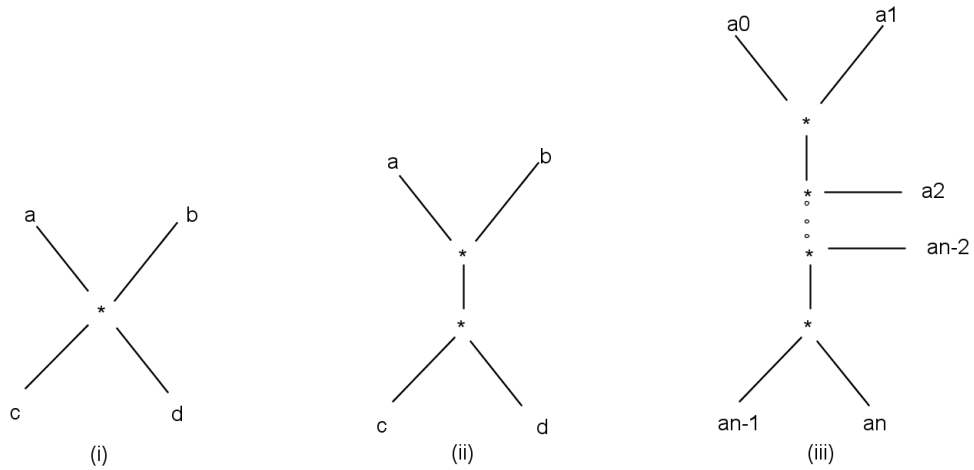


Figura 4.1: X -árvore estrela.

Apresentamos de seguida, alguns exemplos que ilustram a determinação do rank de um lasso \mathcal{L} , associado a uma X -árvore.

Exemplo 16. Consideremos a X -árvore representada na Figura 4.2.

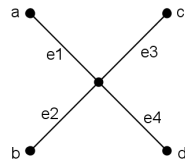


Figura 4.2: X -árvore estrela.

Seja \mathcal{L} um lasso associado a esta X -árvore. Queremos determinar $\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = \dim(\langle \lambda_{xy} : xy \in \mathcal{L} \rangle)$.

Neste caso, $|E| = 4$ e vamos considerar $\mathcal{L} = \{ab, cd, cb, ad\}$. Para determinar o rank de \mathcal{L} , vamos considerar a matriz $M_{\mathcal{L}}$.

Por exemplo, a λ_{ab} está associado um vetor de \mathbb{R}^4 , em que cada entrada desse vetor corresponde à distância da folha a à folha b relativamente a cada uma das funções w_{e_i} . As quatro funções, λ_{xy} com $xy \in \mathcal{L}$, correspondem aos seguintes vetores de \mathbb{R}^4 :

$$\lambda_{ab} \mapsto (\lambda_{ab}(w_{e_1}) \ \lambda_{ab}(w_{e_2}) \ \lambda_{ab}(w_{e_3}) \ \lambda_{ab}(w_{e_4})) = (1 \ 1 \ 0 \ 0)$$

$$\lambda_{cd} \mapsto (\lambda_{cd}(w_{e_1}) \ \lambda_{cd}(w_{e_2}) \ \lambda_{cd}(w_{e_3}) \ \lambda_{cd}(w_{e_4})) = (0 \ 0 \ 1 \ 1)$$

$$\lambda_{cb} \mapsto (\lambda_{cb}(w_{e_1}) \ \lambda_{cb}(w_{e_2}) \ \lambda_{cb}(w_{e_3}) \ \lambda_{cb}(w_{e_4})) = (0 \ 1 \ 1 \ 0)$$

$$\lambda_{ad} \mapsto (\lambda_{ad}(w_{e_1}) \ \lambda_{ad}(w_{e_2}) \ \lambda_{ad}(w_{e_3}) \ \lambda_{ad}(w_{e_4})) = (1 \ 0 \ 0 \ 1)$$

Assim, a matriz é:

$$M_{\mathcal{L}} = \begin{pmatrix} \lambda_{ab} \\ \lambda_{cd} \\ \lambda_{cb} \\ \lambda_{ad} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Como o determinante da matriz $M_{\mathcal{L}}$ é zero então as quatro funções são linearmente dependentes, o que quer dizer que $\dim(< \mathcal{L} >) \leq 3$.

Como $\det \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} = -1 \neq 0$, então a característica da matriz $M_{\mathcal{L}}$ é igual a três,

o que quer dizer que três linhas da matriz $M_{\mathcal{L}}$ são linearmente independentes, ou seja, três dos quatro vetores são linearmente independentes, então a dimensão do espaço vetorial gerado por estas funções é igual a três.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(< \mathcal{L} >) = 3}$$

Consideremos agora: $\mathcal{L} = \{ab, bd, cb, ac\}$.

$$\lambda_{ab} \mapsto (\lambda_{ab}(w_{e_1}) \lambda_{ab}(w_{e_2}) \lambda_{ab}(w_{e_3}) \lambda_{ab}(w_{e_4})) = (1 \ 1 \ 0 \ 0)$$

$$\lambda_{bd} \mapsto (\lambda_{bd}(w_{e_1}) \lambda_{bd}(w_{e_2}) \lambda_{bd}(w_{e_3}) \lambda_{bd}(w_{e_4})) = (0 \ 1 \ 0 \ 1)$$

$$\lambda_{cb} \mapsto (\lambda_{cb}(w_{e_1}) \lambda_{cb}(w_{e_2}) \lambda_{cb}(w_{e_3}) \lambda_{cb}(w_{e_4})) = (0 \ 1 \ 1 \ 0)$$

$$\lambda_{ac} \mapsto (\lambda_{ac}(w_{e_1}) \lambda_{ac}(w_{e_2}) \lambda_{ac}(w_{e_3}) \lambda_{ac}(w_{e_4})) = (1 \ 0 \ 1 \ 0)$$

e a seguinte matriz:

$$M_{\mathcal{L}} = \begin{pmatrix} \lambda_{ab} \\ \lambda_{bd} \\ \lambda_{cb} \\ \lambda_{ac} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Como $\det(M_{\mathcal{L}}) = 2 \neq 0$, então as quatro funções são linearmente independentes, o que quer dizer que a dimensão do espaço gerado por elas é igual a 4.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(< \mathcal{L} >) = 4}$$

Assim sendo, \mathcal{L} é um lasso com $\text{rank}(\mathcal{L}) = 4$ e $|\mathcal{L}| = 4 = |E| = \dim \widehat{\mathbb{R}^E}$. Portanto, \mathcal{L} é uma base do matroide $\mathcal{M}(T)$.

Observe-se que \mathcal{L} também é um *edge-weight* lasso. Dada uma função peso $w \in \mathbb{R}^E$, que atribui um valor $w(e_i) \in \mathbb{R}$ a cada aresta $e_i \in E$, seja $D_w(x, y)$ a distância entre o vértice x e y relativamente a w . Supondo que as distâncias (relativamente a w) entre os pares de folhas em \mathcal{L} são conhecidas, temos

$$\begin{cases} D_w(a, b) = x_0 \\ D_w(b, d) = x_1 \\ D_w(c, b) = x_2 \\ D_w(a, c) = x_3 \end{cases}, x_0, x_1, x_2, x_3 \in \mathbb{R} \Leftrightarrow \begin{cases} w(e_1) + w(e_2) = x_0 \\ w(e_2) + w(e_4) = x_1 \\ w(e_2) + w(e_3) = x_2 \\ w(e_1) + w(e_3) = x_3 \end{cases}.$$

Observe-se que a matriz formada pelos coeficientes do sistema é igual a $M_{\mathcal{L}}$. O determinante da matriz é diferente de zero, então o sistema é possível e determinado, logo \mathcal{L} é um *edge-weight* lasso, já que determina w .

Na Figura 4.3 está representado o grafo $\Gamma(\mathcal{L})$ associado ao lasso \mathcal{L} .

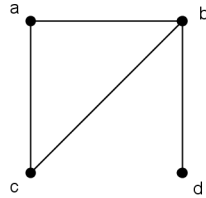


Figura 4.3: Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bd, cb, ac\}$.

Exemplo 17. Consideremos a X -árvore representada na Figura 4.4. Nos mesmos padrões que o exemplo anterior, queremos determinar, para um certo lasso \mathcal{L} , o seu rank.

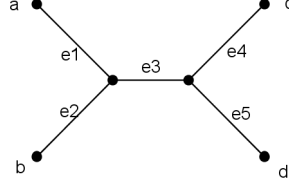


Figura 4.4: X -árvore quarteto.

Nesta árvore $|E| = 5$. Seja $\mathcal{L} = \{ab, cd\}$ um lasso associado a esta X -árvore. Como λ_{ab} e λ_{cd} são duas funções que não se podem escrever à custa uma da outra, então são linearmente independentes, logo $\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 2$.

Consideremos agora $\mathcal{L} = \{ab, cd, ad, bc\}$.

$$\lambda_{ab} \mapsto (\lambda_{ab}(w_{e_1}) \lambda_{ab}(w_{e_2}) \lambda_{ab}(w_{e_3}) \lambda_{ab}(w_{e_4}) \lambda_{ab}(w_{e_5})) = (1 \ 1 \ 0 \ 0 \ 0)$$

$$\lambda_{cd} \mapsto (\lambda_{cd}(w_{e_1}) \lambda_{cd}(w_{e_2}) \lambda_{cd}(w_{e_3}) \lambda_{cd}(w_{e_4}) \lambda_{cd}(w_{e_5})) = (0 \ 0 \ 0 \ 1 \ 1)$$

$$\lambda_{ad} \mapsto (\lambda_{ad}(w_{e_1}) \lambda_{ad}(w_{e_2}) \lambda_{ad}(w_{e_3}) \lambda_{ad}(w_{e_4}) \lambda_{ad}(w_{e_5})) = (1 \ 0 \ 1 \ 0 \ 1)$$

$$\lambda_{bc} \mapsto (\lambda_{bc}(w_{e_1}) \lambda_{bc}(w_{e_2}) \lambda_{bc}(w_{e_3}) \lambda_{bc}(w_{e_4}) \lambda_{bc}(w_{e_5})) = (0 \ 1 \ 1 \ 1 \ 0)$$

Uma forma de calcular a dimensão do subespaço gerado pelas quatro funções é verificar se existem x, y, z e $w \in \mathbb{R}$ tal que:

$$x\lambda_{ab} + y\lambda_{cd} + z\lambda_{ad} + w\lambda_{bc} = 0.$$

A solução $x = y = z = w = 0$ é chamada a solução trivial da equação. Se essa solução for única, então as funções são linearmente independentes. Caso a solução trivial não seja a única, então as funções são linearmente dependentes.

A dimensão do subespaço gerado pelas quatro funções é menor ou igual a quatro. Para verificar se as quatro funções são linearmente dependentes ou independentes, resolvemos o sistema:

$$\begin{cases} x\lambda_{ab}(w_{e_1}) + y\lambda_{cd}(w_{e_1}) + z\lambda_{ad}(w_{e_1}) + w\lambda_{bc}(w_{e_1}) = 0 \\ x\lambda_{ab}(w_{e_2}) + y\lambda_{cd}(w_{e_2}) + z\lambda_{ad}(w_{e_2}) + w\lambda_{bc}(w_{e_2}) = 0 \\ x\lambda_{ab}(w_{e_3}) + y\lambda_{cd}(w_{e_3}) + z\lambda_{ad}(w_{e_3}) + w\lambda_{bc}(w_{e_3}) = 0 \\ x\lambda_{ab}(w_{e_4}) + y\lambda_{cd}(w_{e_4}) + z\lambda_{ad}(w_{e_4}) + w\lambda_{bc}(w_{e_4}) = 0 \\ x\lambda_{ab}(w_{e_5}) + y\lambda_{cd}(w_{e_5}) + z\lambda_{ad}(w_{e_5}) + w\lambda_{bc}(w_{e_5}) = 0 \end{cases} \Leftrightarrow \begin{cases} x + z = 0 \\ x + w = 0 \\ z + w = 0 \\ y + w = 0 \\ y + z = 0 \end{cases} \Leftrightarrow \begin{cases} z = -x \\ w = -x \\ - \\ y = -w \\ z = -y \end{cases} \Leftrightarrow \begin{cases} z = 0 \\ w = 0 \\ -x - x = 0 \Rightarrow x = 0 \\ y = 0 \\ z = 0 \end{cases}$$

$$\boxed{\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 4}$$

Este sistema só tem uma solução, a solução trivial, o que quer dizer que as quatro funções são linearmente independentes e a dimensão do subespaço gerado por elas é igual à sua cardinalidade.

Uma base do matroide associado a esta X -árvore é um lasso \mathcal{L} , com cardinalidade igual a $|E| = 5$ e $\text{rank}(\mathcal{L}) = 5$.

Vamos verificar se o lasso $\mathcal{L} = \{ab, cd, ad, bc, bd\}$ tem as propriedades referidas acima.

Com o intuito de calcular a dimensão do subespaço gerado pelas cinco funções, vamos considerar a matriz formada pelos vetores λ_{xy} , com $xy \in \mathcal{L}$:

$$M_{\mathcal{L}} = \begin{pmatrix} \lambda_{ab} \\ \lambda_{cd} \\ \lambda_{ad} \\ \lambda_{bc} \\ \lambda_{bd} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Como $\det(M_{\mathcal{L}}) = -4 \neq 0$, então as cinco funções são linearmente independentes, logo a dimensão do subespaço gerado pelas mesmas é igual à sua cardinalidade.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 5}$$

Portanto \mathcal{L} é uma base do matroide $\mathcal{M}(T)$.

Observe-se que \mathcal{L} é um *edge-weight* lasso. Dada uma função peso $w \in \mathbb{R}^E$, que atribui um valor $w(e_i) \in \mathbb{R}$ a cada aresta $e_i \in E$, seja $D_w(x, y)$ a distância entre o vértice x e y relativamente a w . Supondo que as distâncias (relativamente a w) entre os pares de folhas em \mathcal{L} são conhecidas, temos

$$\begin{cases} D_w(a, b) = x_0 \\ D_w(c, d) = x_1 \\ D_w(a, d) = x_2 \\ D_w(b, c) = x_3 \\ D_w(b, d) = x_4 \end{cases} \Leftrightarrow \begin{cases} w(e_1) + w(e_2) = x_0 \\ w(e_4) + w(e_5) = x_1 \\ w(e_1) + w(e_3) + w(e_5) = x_2 \\ w(e_2) + w(e_3) + w(e_4) = x_3 \\ w(e_2) + w(e_3) + w(e_5) = x_4 \end{cases}$$

Observe-se que a matriz formada pelos coeficientes do sistema é igual a $M_{\mathcal{L}}$. O determinante da matriz é diferente de zero, então o sistema é possível e determinado, logo \mathcal{L} é um *edge-weight* lasso, já que determina w .

Na Figura 4.5 está representado o grafo $\Gamma(\mathcal{L})$ associado ao lasso \mathcal{L} .

Exemplo 18. Consideremos a X -árvore representada na Figura 4.6. Seguindo o modelo dos exemplos anteriores, queremos determinar, para um certo lasso \mathcal{L} , o seu rank.

Nesta X -árvore $|E| = 7$. Seja $\mathcal{L} = \{ab, be, de, cd\}$ um lasso associado a esta X -árvore.

$$\lambda_{ab} \mapsto (\lambda_{ab}(w_{e_1}) \lambda_{ab}(w_{e_2}) \lambda_{ab}(w_{e_3}) \lambda_{ab}(w_{e_4}) \lambda_{ab}(w_{e_5}) \lambda_{ab}(w_{e_6}) \lambda_{ab}(w_{e_7})) = (1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)$$

$$\lambda_{be} \mapsto (\lambda_{be}(w_{e_1}) \lambda_{be}(w_{e_2}) \lambda_{be}(w_{e_3}) \lambda_{be}(w_{e_4}) \lambda_{be}(w_{e_5}) \lambda_{be}(w_{e_6}) \lambda_{be}(w_{e_7})) = (0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0)$$

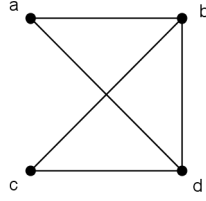


Figura 4.5: Grafo $\Gamma(\mathcal{L})$ associado ao laço $\mathcal{L} = \{ab, cd, ad, bc, bd\}$.

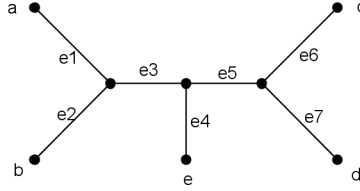


Figura 4.6: X -árvore lagarta.

$$\lambda_{de} \mapsto (\lambda_{de}(w_{e_1}) \ \lambda_{de}(w_{e_2}) \ \lambda_{de}(w_{e_3}) \ \lambda_{de}(w_{e_4}) \ \lambda_{de}(w_{e_5}) \ \lambda_{de}(w_{e_6}) \ \lambda_{de}(w_{e_7})) = (0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1)$$

$$\lambda_{cd} \mapsto (\lambda_{cd}(w_{e_1}) \ \lambda_{cd}(w_{e_2}) \ \lambda_{cd}(w_{e_3}) \ \lambda_{cd}(w_{e_4}) \ \lambda_{cd}(w_{e_5}) \ \lambda_{cd}(w_{e_6}) \ \lambda_{cd}(w_{e_7})) = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1)$$

Uma forma de calcular a dimensão do subespaço gerado pelas quatro funções é verificar se existem x, y, z e $w \in \mathbb{R}$ tal que:

$$x\lambda_{ab} + y\lambda_{be} + z\lambda_{de} + w\lambda_{cd} = 0.$$

Resolvendo o sistema:

$$\begin{cases} x\lambda_{ab}(w_{e_1}) + y\lambda_{be}(w_{e_1}) + z\lambda_{de}(w_{e_1}) + w\lambda_{cd}(w_{e_1}) = 0 \\ x\lambda_{ab}(w_{e_2}) + y\lambda_{be}(w_{e_2}) + z\lambda_{de}(w_{e_2}) + w\lambda_{cd}(w_{e_2}) = 0 \\ x\lambda_{ab}(w_{e_3}) + y\lambda_{be}(w_{e_3}) + z\lambda_{de}(w_{e_3}) + w\lambda_{cd}(w_{e_3}) = 0 \\ x\lambda_{ab}(w_{e_4}) + y\lambda_{be}(w_{e_4}) + z\lambda_{de}(w_{e_4}) + w\lambda_{cd}(w_{e_4}) = 0 \Leftrightarrow \\ x\lambda_{ab}(w_{e_5}) + y\lambda_{be}(w_{e_5}) + z\lambda_{de}(w_{e_5}) + w\lambda_{cd}(w_{e_5}) = 0 \\ x\lambda_{ab}(w_{e_6}) + y\lambda_{be}(w_{e_6}) + z\lambda_{de}(w_{e_6}) + w\lambda_{cd}(w_{e_6}) = 0 \\ x\lambda_{ab}(w_{e_7}) + y\lambda_{be}(w_{e_7}) + z\lambda_{de}(w_{e_7}) + w\lambda_{cd}(w_{e_7}) = 0 \end{cases}$$

$$\begin{cases} x = 0 \\ x + y = 0 \\ y = 0 \\ y + z = 0 \\ z = 0 \\ w = 0 \\ z + w = 0 \end{cases} \Leftrightarrow \begin{cases} x = 0 \\ y = 0 \\ z = 0 \\ w = 0 \end{cases}.$$

Como a única solução do sistema é a solução trivial, então as quatro funções são linearmente independentes, o que quer dizer que a dimensão do subespaço gerado por elas é igual à sua cardinalidade.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 4}$$

Utilizando o mesmo processo, determinamos que:

$$\mathcal{L} = \{ab, be, de, cd, ce\} : \text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 5$$

$$\mathcal{L} = \{ab, be, de, cd, ce, ae\} : \text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 6$$

Com o intuito de determinar uma base do matroide associado à X -árvore, consideramos o lasso $\mathcal{L} = \{ab, bc, cd, ad, be, ce, bd\}$, com $|\mathcal{L}| = |E| = 7$.

Para calcular o rank do lasso, a dimensão do subespaço gerado pelas sete funções, vamos considerar a matriz formada pelos vetores λ_{xy} , com $xy \in \mathcal{L}$:

$$M_{\mathcal{L}} = \begin{pmatrix} \lambda_{ab} \\ \lambda_{bc} \\ \lambda_{cd} \\ \lambda_{ad} \\ \lambda_{be} \\ \lambda_{ce} \\ \lambda_{bd} \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Como o $\det(M_{\mathcal{L}}) = -8 \neq 0$, então as sete funções são linearmente independentes, o que quer dizer que a dimensão do subespaço gerado pelas mesmas é igual à sua cardinalidade.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 7}$$

Observe-se que \mathcal{L} é um *edge-weight* lasso. Dada uma função peso $w \in \mathbb{R}^E$, que atribui um valor $w(e_i) \in \mathbb{R}$ a cada aresta $e_i \in E$, seja $D_w(x, y)$ a distância entre o vértice x e y relativamente a w . Supondo que as distâncias (relativamente a w) entre os pares de folhas em \mathcal{L} são conhecidas, temos

$$\begin{cases} D_w(a, b) = x_0 \\ D_w(b, e) = x_1 \\ D_w(d, e) = x_2 \\ D_w(c, d) = x_3 \\ D_w(c, e) = x_4 \\ D_w(a, e) = x_5 \\ D_w(b, d) = x_6 \end{cases} \Leftrightarrow \begin{cases} w(e_1) + w(e_2) = x_0 \\ w(e_2) + w(e_3) + w(e_5) + w(e_6) = x_1 \\ w(e_6) + w(e_7) = x_2 \\ w(e_1) + w(e_3) + w(e_5) + w(e_7) = x_3 \\ w(e_2) + w(e_3) + w(e_4) = x_4 \\ w(e_4) + w(e_5) + w(e_6) = x_5 \\ w(e_2) + w(e_3) + w(e_5) + w(e_7) = x_6 \end{cases}$$

Observe-se que a matriz formada pelos coeficientes do sistema é igual a $M_{\mathcal{L}}$. O determinante da matriz é diferente de zero, então o sistema é possível e determinado, logo \mathcal{L} é um *edge-weight* lasso, já que determina w .

Na Figura 4.7 está representada o grafo $\Gamma(\mathcal{L})$ associado ao lasso \mathcal{L} .

Exemplo 19. Consideremos a X -árvore representada na Figura 4.8. Seguindo o modelo dos exemplos anteriores, queremos determinar, para um certo lasso \mathcal{L} , o seu rank.

Com o intuito de determinar uma base do matroide associado à X -árvore, consideramos o lasso $\mathcal{L} = \{ab, bc, cd, ad, ac, ae, bf, cf, de\}$, $|\mathcal{L}| = |E(T)|$.

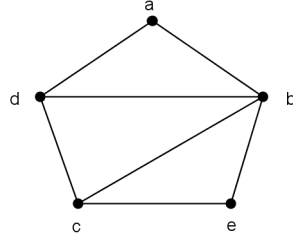


Figura 4.7: Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bc, cd, ad, be, ce, bd\}$.

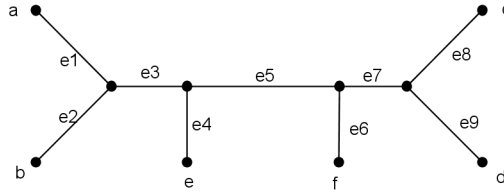


Figura 4.8: X -árvore lagarta.

Para calcular o rank do lasso, a dimensão do subespaço gerado pelas sete funções, vamos considerar a matriz formada pelos vetores $\lambda_{xy}, \forall xy \in \mathcal{L}$:

$$M_{\mathcal{L}} = \begin{pmatrix} \lambda_{ab} \\ \lambda_{bc} \\ \lambda_{cd} \\ \lambda_{ad} \\ \lambda_{ac} \\ \lambda_{ae} \\ \lambda_{bf} \\ \lambda_{cf} \\ \lambda_{de} \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Como o $\det(M_{\mathcal{L}}) = 16 \neq 0$, então as nove funções são linearmente independentes, o que quer dizer que a dimensão do subespaço gerado pelas mesmas é igual à sua cardinalidade.

$$\boxed{\text{rank}(\mathcal{L}) = \dim(\langle \mathcal{L} \rangle) = 9}$$

Na Figura 4.9 está representada o grafo $\Gamma(\mathcal{L})$ associado ao lasso \mathcal{L} .

Teorema 4.2. *Seja $T = (V, E)$ uma X -árvore e $\mathcal{L} \subseteq \binom{X}{2}$. As seguintes condições são equivalentes:*

- (a) \mathcal{L} é um edge-weight lasso;
- (b) Dado $w_0 \in \mathbb{R}^E$, se $\lambda_{xy}^T(w_0) = 0$ para toda a corda $xy \in \mathcal{L}$, então $w_0 = 0$;
- (c) $\dim\langle \mathcal{L} \rangle = |E|$.

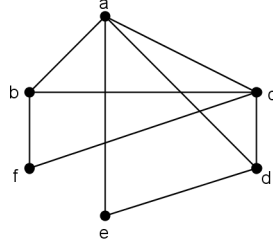


Figura 4.9: Grafo $\Gamma(\mathcal{L})$ associado ao lasso $\mathcal{L} = \{ab, bc, cd, ad, ac, ae, bf, cf, de\}$.

Em particular, se \mathcal{L} é um *edge-weight lasso*, então $\cup \mathcal{L} = X$ e $|\mathcal{L}| \geq |E|$. Se \mathcal{L} é um *edge-weight lasso minimal* então $|\mathcal{L}| = |E|$.

Demonstração. Prova de (a) \implies (b): Suponhamos que \mathcal{L} é um *edge-weight lasso* e que $w_0 \in \mathbb{R}^E$ é tal que $\lambda_{xy}^T(w_0) = 0$ para toda a corda $xy \in \mathcal{L}$. Seja $\omega : E \rightarrow \mathbb{R}_{>0}$ uma função de peso nas arestas de T (por exemplo, $w = \sum_{e \in E} w_e$). Então existe $\epsilon \neq 0$ tal que $w + \epsilon w_0$ toma valores em $\mathbb{R}_{>0}$. Além disso, $\lambda_{xy}^T(w + \epsilon w_0) = \lambda_{xy}^T(w) + \epsilon \lambda_{xy}^T(w_0) = \lambda_{xy}^T(w)$, logo $(T, w) \stackrel{\mathcal{L}}{\equiv} (T, w + \epsilon w_0)$. Como \mathcal{L} é um *edge-weight lasso*, resulta que $w = w + \epsilon w_0$ e assim $w_0 = 0$ pois $\epsilon \neq 0$.

Suponhamos ainda, por redução ao absurdo, que existe $z \in X \setminus \cup \mathcal{L}$, e seja $f \in E$ a única aresta incidente com z . Então para todo o $xy \in \mathcal{L}$, $\lambda_{xy}^T(w_f) = 0$, já que $f \notin E_T(xy)$. Pelo que acabámos de mostrar, terá de ser $w_f = 0$, o que é absurdo. Logo $\cup \mathcal{L} = X$.

Prova de (b) \implies (a): Suponhamos que (b) se verifica. Sejam $w, w' : E \rightarrow \mathbb{R}_{\geq 0}$ duas funções de peso próprias (isto é, w e w' tomam valores positivos nas arestas interiores de T) e tais que $(T, w) \stackrel{\mathcal{L}}{\equiv} (T, w')$. Então $\lambda_{xy}^T(w) = \lambda_{xy}^T(w')$ para todo o $xy \in \mathcal{L}$, e assim $w - w'$ anula todas as funções λ_{xy}^T . Conclui-se portanto de (b) que $w - w' = 0$, o que mostra que \mathcal{L} é um *edge-weight lasso*.

Seja $\langle \mathcal{L} \rangle$ o subespaço de $\widehat{\mathbb{R}^E}$ gerado por $\{\lambda_{xy}^T : xy \in \mathcal{L}\}$. Como $\dim \widehat{\mathbb{R}^E} = |E|$, temos $\dim \langle \mathcal{L} \rangle \leq |E|$. Considere-se a aplicação linear

$$\Phi : \mathbb{R}^E \longrightarrow \widehat{\langle \mathcal{L} \rangle}, \quad w \mapsto \Phi_w,$$

com $\Phi_w(f) = f(w)$, para $f \in \langle \mathcal{L} \rangle$. Temos

$$\begin{aligned} w \in \ker \Phi &\Leftrightarrow \Phi_w(f) = 0 \quad \forall f \in \langle \mathcal{L} \rangle \\ &\Leftrightarrow \Phi_w(\lambda_{xy}^T) = 0 \quad \forall xy \in \mathcal{L} \\ &\Leftrightarrow \lambda_{xy}^T(w) = 0 \quad \forall xy \in \mathcal{L}. \end{aligned}$$

Logo, (b) verifica-se se e só se Φ é injetiva. Assim, assumindo (b) temos

$$|E| = \dim \mathbb{R}^E = \dim \text{im} \Phi \leq \dim \widehat{\langle \mathcal{L} \rangle} = \dim \langle \mathcal{L} \rangle \leq |E|,$$

o que implica que $\dim \langle \mathcal{L} \rangle = |E|$.

Resta verificar que (c) implica (b). Suponhamos então que $\dim \langle \mathcal{L} \rangle = |E|$, e portanto que $\langle \mathcal{L} \rangle = \widehat{\mathbb{R}^E}$. Se $0 \neq w \in \mathbb{R}^E$, então existe $\hat{w} \in \widehat{\mathbb{R}^E}$ tal que $\hat{w}(w) \neq 0$. Como $\langle \mathcal{L} \rangle = \widehat{\mathbb{R}^E}$,

existem escalares $c_{xy} \in \mathbb{R}$, para cada $xy \in \mathcal{L}$, tais que $\hat{w} = \sum_{xy \in \mathcal{L}} c_{xy} \lambda_{xy}^T$ e

$$0 \neq \hat{w}(w) = \sum_{xy \in \mathcal{L}} c_{xy} \lambda_{xy}^T(w).$$

Logo existe $xy \in \mathcal{L}$ tal que $\lambda_{xy}^T(w) \neq 0$, o que prova (b).

Continuando a supor (c), temos que, como $\{\lambda_{xy}^T : xy \in \mathcal{L}\}$ gera $\langle \mathcal{L} \rangle$, por definição decorre que $|\mathcal{L}| \geq |\{\lambda_{xy}^T : xy \in \mathcal{L}\}| \geq \dim \langle \mathcal{L} \rangle = |E|$. Se \mathcal{L} é um *edge-weight* lasso e $|\mathcal{L}| > |E|$ então as funções λ_{xy}^T , com $xy \in \mathcal{L}$, não são linearmente independentes, e existe $xy \in \mathcal{L}$ tal que $\mathcal{L} \setminus \{xy\}$ é um *edge-weight* lasso, o que prova que \mathcal{L} não é minimal. \square

4.2 Árvores estrela

Nesta secção vamos estudar um tipo de X -árvores mais simples, as X -árvores estrela (ver Definição 24 e a Figura 4.1).

A proposição seguinte caracteriza o matroide $\mathcal{M}(T^*)$ onde T^* é uma X -árvore estrela.

Proposição 1 ([4, Proposição 3.1]). Seja T^* uma X -árvore estrela com $|X| \geq 3$. As seguintes afirmações são válidas para o matroide $\mathcal{M}(T^*)$:

- A coleção dos conjuntos geradores deste matroide, $\mathcal{G}(T^*)$, isto é, a coleção de todos os *edge-weight* lassos, coincide com a coleção de todos os subconjuntos \mathcal{L} de $\binom{X}{2}$, com cardinalidade maior ou igual ao número de arestas de T^* , para os quais o seu grafo associado, $\Gamma(\mathcal{L})$, é fortemente não bipartido.
- A coleção das bases, $\mathcal{B}(T^*)$, isto é, a coleção de todos os *edge-weight* lassos com cardinalidade mínima, coincide com a coleção de todos os subconjuntos minimais \mathcal{L} de $\binom{X}{2}$, para os quais o seu grafo associado, $\Gamma(\mathcal{L})$, é fortemente não bipartido, isto é, cada componente conexa de $\Gamma(\mathcal{L})$ tem exatamente um ciclo de comprimento ímpar.
- A coleção dos independentes, $\mathcal{I}(T^*)$, coincide com coleção de subconjuntos \mathcal{L} de $\binom{X}{2}$, para os quais cada componente conexa do seu grafo associado, $\Gamma(\mathcal{L})$, ou é uma árvore ou contém exatamente um ciclo de comprimento ímpar.
- A coleção dos circuitos, $\mathcal{C}(T^*)$, coincide com a coleção dos subconjuntos \mathcal{L} de $\binom{X}{2}$, para os quais o grafo $\Gamma(\mathcal{L})$, forma ou um único circuito de comprimento par, ou um par de circuitos, de comprimento ímpar, juntamente com um caminho simples de ligação entre eles, de tal modo que os dois ciclos ou são disjuntos ou têm exatamente um vértice em comum.

Vamos ver alguns exemplos que ilustram a proposição.

Nos exemplos que se seguem vamos denotar as arestas do grafo $\Gamma(\mathcal{L})$ por **cordas**, para que não se confundam as arestas das X -árvores com as arestas deste grafo.

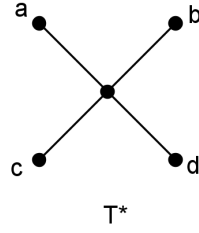


Figura 4.10: X -árvore estrela.

Exemplo 20. Como no Exemplo 16, vamos considerar uma X -árvore estrela, representada na Figura 4.10, com conjunto de folhas $X := \{a, b, c, d\}$.

O grafo $\Gamma(\mathcal{L})$ é um grafo com quatro vértices (a, b, c, d) e, no máximo, seis cordas, porque $|\binom{X}{2}| = \binom{4}{2} = 6$.

Como vimos na Proposição 1, a **coleção dos geradores** é a coleção de todos os subconjuntos \mathcal{L} de $\binom{X}{2}$, com cardinalidade maior ou igual a $|E(T^*)| = 4$, para os quais o grafo associado, $\Gamma(\mathcal{L})$, é fortemente não bipartido.

Podemos ter grafos com quatro cordas, e sabemos que existem seis possibilidades de cordas, logo existem $\binom{6}{4} = 15$ possibilidades de grafos com quatro cordas, representados na Figura 4.11.

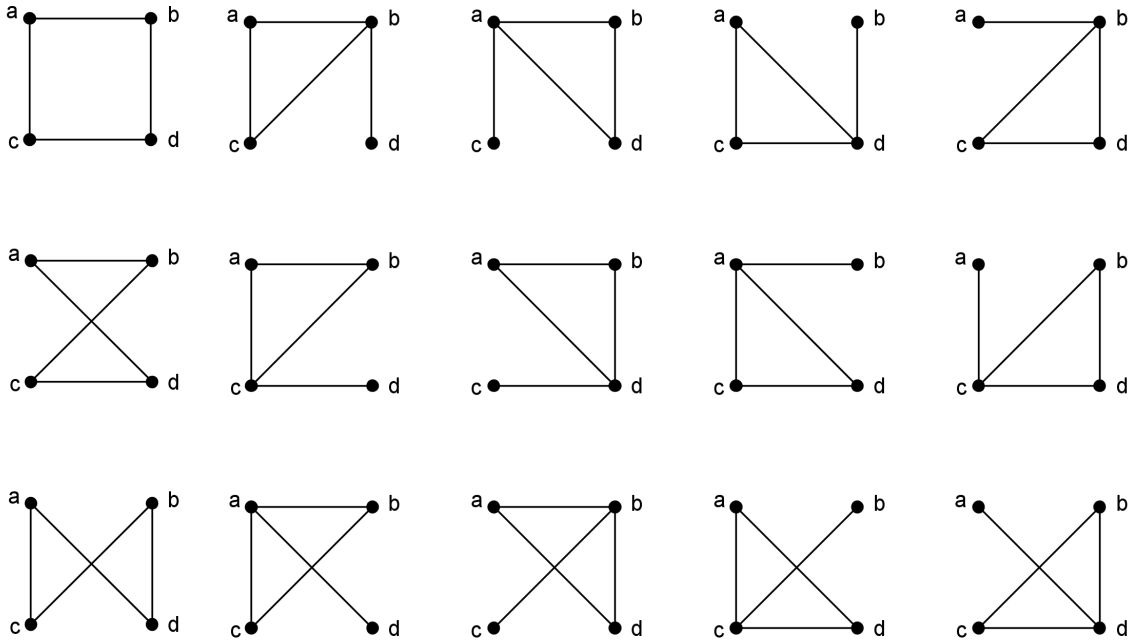


Figura 4.11: Grafos $\Gamma(\mathcal{L})$ associados aos lassos \mathcal{L} com quatro cordas.

Como o grafo tem que ser fortemente não bipartido, ou seja, todas as suas componentes conexas têm que ser não bipartidas, os lassos associados aos grafos da primeira coluna da Figura 4.11 não são conjuntos geradores do matroide associado a T^* , mas os restantes são. Logo, a família de geradores $\mathcal{G}(T^*)$ com quatro elementos, é

constituída por 12 subconjuntos \mathcal{L} correspondentes aos grafos da Figura 4.11, exceto os da primeira coluna.

Podemos ter também conjuntos de cardinalidade cinco e seis, para os quais existem $\binom{6}{5} = 6$ e $\binom{6}{6} = 1$ possibilidades, respetivamente, e nesses casos $\Gamma(\mathcal{L})$ é fortemente não bipartido, pois é conexo e contém um triângulo.

Logo, a família de geradores contém 19 subconjuntos \mathcal{L} .

Por outro lado, a **coleção das bases** são todos os subconjuntos \mathcal{L} , com cardinalidade mínima, para o quais o grafo associado, $\Gamma(\mathcal{L})$, é fortemente não bipartido e tem pelo menos um ciclo de comprimento ímpar.

Sabemos que os lassos \mathcal{L} que pertencem ao conjuntos das bases são subconjuntos geradores com cardinalidade mínima, ou seja, cardinalidade igual a quatro. Logo, essas bases correspondem aos grafos $\Gamma(\mathcal{L})$ representados na Figura 4.12.

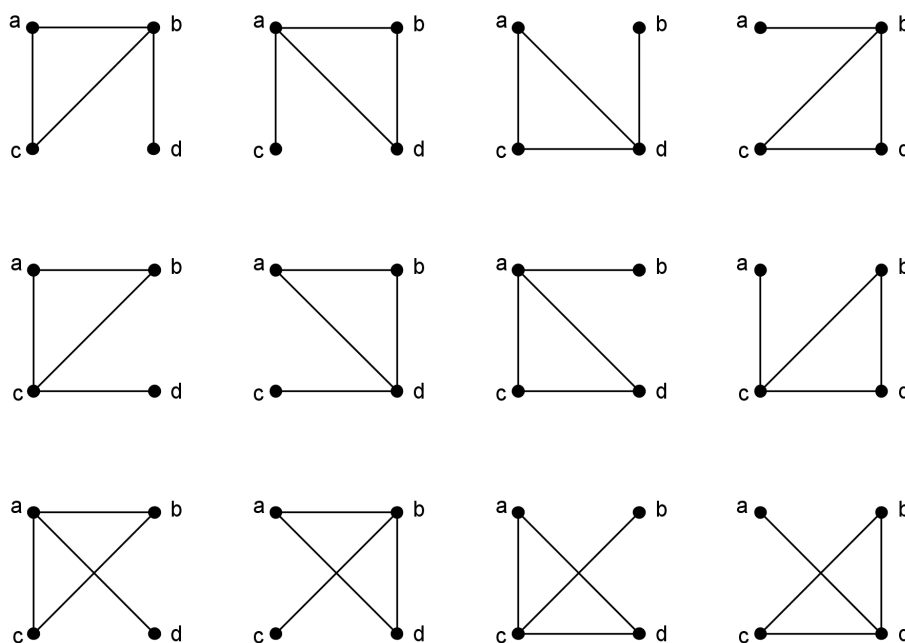


Figura 4.12: Grafos $\Gamma(\mathcal{L})$ associados aos subconjuntos \mathcal{L} que são bases da X -árvore da Figura 4.10.

A **coleção dos independentes**, coincide com a coleção dos subconjuntos \mathcal{L} para os quais cada componente conexa do seu grafo associado, $\Gamma(\mathcal{L})$, ou é uma árvore ou contém exatamente um ciclo de comprimento ímpar. O grafo vazio com quatro vértices corresponde ao independente $\mathcal{L} = \emptyset$ de rank 0.

Qualquer corda isolada é um independente, por isso existem seis hipóteses para este caso.

No caso em que temos duas cordas, existem $\binom{6}{2} = 15$ hipóteses, pois quaisquer duas cordas neste caso formam uma floresta.

Um ciclo de comprimento ímpar tem neste caso comprimento três. Os grafos constituídos por três cordas, ou formam um ciclo de comprimento três ou uma floresta, por isso existem $\binom{6}{3} = 20$ hipóteses para este caso.

As bases são os independentes com rank quatro, para as quais existem 12 hipóteses já indicadas.

Não há independentes com rank superior a quatro, logo já foram descritos todos os independentes.

Portanto, a família de independentes do matroide é constituída por $1+6+15+20+12 = 54$ conjuntos de independentes.

A **coleção dos circuitos**, é formada pelos subconjuntos \mathcal{L} para os quais o seu grafo associado, $\Gamma(\mathcal{L})$, ou forma um só ciclo par ou dois ciclos ímpares ligados por um caminho de comprimento maior ou igual a zero (se for maior que zero são disjuntos).

No caso em que $\Gamma(\mathcal{L})$ forma um ciclo de comprimento par, este será um ciclo de comprimento quatro, e há três possibilidades para este ciclo (ver primeira coluna da Figura 4.11).

Não é possível a existência de um par de ciclos de comprimento ímpar nas condições descritas acima. Para verificar que não é possível a existência de um par de ciclos de comprimento ímpar ligados por um caminho de comprimento maior ou igual a zero, vamos pensar no grafo $\Gamma(\mathcal{L})$ como um subgrafo do grafo completo K_4 , ilustrado na Figura 4.13, verificamos facilmente que a existência de tais ciclos é impossível.

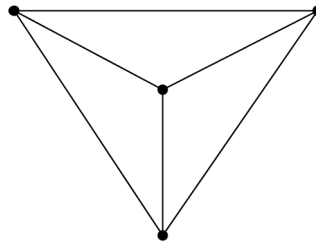


Figura 4.13: Grafo completo K_4 .

Exemplo 21. Vamos considerar uma X -árvore estrela, representada na Figura 4.14, com conjunto de folhas $X := \{a, b, c, d, e\}$.

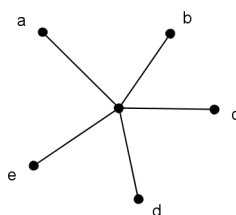


Figura 4.14: X -árvore estrela.

Para qualquer laço \mathcal{L} , o grafo $\Gamma(\mathcal{L})$ é um grafo com cinco vértices e, no máximo, dez cordas, já que $|\binom{X}{2}| = \binom{5}{2} = 10$.

Como no exemplo anterior vamos utilizar a Proposição 1.

Vamos pensar no grafo $\Gamma(\mathcal{L})$ como um subgrafo do grafo completo com cinco vértices, K_5 .

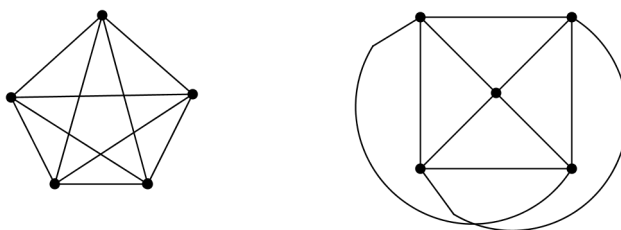


Figura 4.15: Duas representações do grafo completo com cinco vértices, K_5 .

A **coleção dos geradores** é formada pelos subconjuntos \mathcal{L} com cardinal maior ou igual a cinco e tal que o seu grafo associado, $\Gamma(\mathcal{L})$ é fortemente não bipartido.

Neste caso, o grafo $\Gamma(\mathcal{L})$ pode conter 5, 6, 7, 8, 9 ou 10 cordas.

No caso em que contém cinco cordas, existem $\binom{10}{5} = 252$ grafos possíveis, mas temos que ver quais destes é que são fortemente não bipartidos. Note-se que um subgrafo de K_5 com cinco ou mais arestas e sem vértices isolados é necessariamente conexo, logo $\Gamma(\mathcal{L})$ é um grafo conexo e não bipartido.

Pelo Teorema 2.1 sabemos que um grafo é bipartido se e só se não contém ciclos de comprimento ímpar. Em particular, uma árvore é um grafo bipartido.

Os geradores \mathcal{L} de $\mathcal{M}(T)$ com cinco cordas são precisamente as bases de $\mathcal{M}(T)$ e $\Gamma(\mathcal{L})$ é caracterizado por ser um grafo conexo com exatamente um ciclo de comprimento ímpar. Neste caso, o grafo $\Gamma(\mathcal{L})$ ou é constituído por um único ciclo de comprimento cinco ou por um ciclo de comprimento três e duas arestas.

No caso em que o ciclo tem comprimento três temos dez hipóteses, ilustradas na Figura 4.16.

Como o grafo é conexo, os dois vértices restantes têm que formar uma das configurações da Figura 4.17.

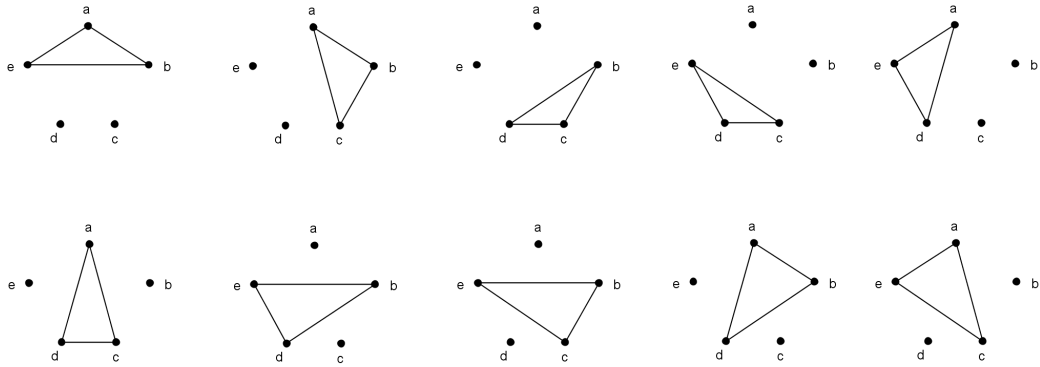


Figura 4.16: Hipóteses dos ciclos de comprimento três.

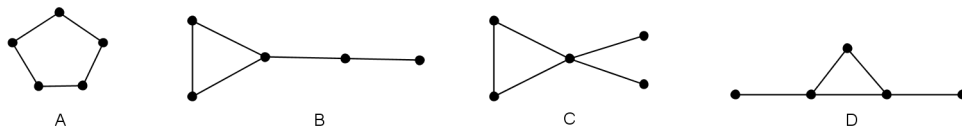


Figura 4.17: Configurações para as bases de $\mathcal{M}(T)$.

Para a hipótese A existem três possibilidades, para a hipótese B existem três e para a hipótese C existem seis. Ou seja, para grafos construídos desta forma temos $10 \times 3 + 10 \times 3 + 10 \times 6 = 120$ hipóteses. No caso de $\Gamma(\mathcal{L})$ ser um ciclo de comprimento cinco temos $\frac{5!}{5 \times 2} = 12$ hipóteses. Assim, no total, $\mathcal{M}(T)$ tem $120 + 12 = 132$ bases e 132 geradores com cinco elementos.

Se \mathcal{L} é um gerador com seis elementos, então $\Gamma(\mathcal{L})$ é conexo, tem algum ciclo de comprimentos ímpar e tem seis cordas. Há $\binom{10}{6} = 210$ subgrafos de K_5 com seis cordas. Destes, todos são conexos exceto os cinco que têm uma componente conexa isomorfa a K_4 e um vértice isolado. De entre os que são conexos, todos têm algum ciclo de comprimento ímpar exceto os $\binom{5}{2} = 10$ que correspondem à configuração representada na Figura 4.18.

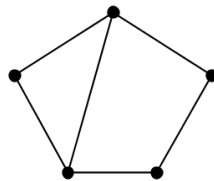


Figura 4.18: Configuração para o grafo $\Gamma(\mathcal{L})$ que tem um ciclo de comprimento par.

Logo, há $210 - 5 - 10 = 195$ subgrafos conexos de K_5 com seis cordas e que não são bipartidos, o que significa que $\mathcal{M}(T)$ tem 195 geradores com seis elementos.

Com sete ou mais cordas todos os subgrafos de K_5 são necessariamente conexos e têm pelo menos algum ciclo de comprimento ímpar, logo há $\binom{10}{7} = 120$, $\binom{10}{8} = 45$, $\binom{10}{9} = 10$, $\binom{10}{10} = 1$ geradores de $\mathcal{M}(T)$ com sete, oito, novo ou dez cordas, respetivamente.

Ou seja, existem $132 + 195 + 120 + 45 + 10 + 1 = 503$ geradores de $\mathcal{M}(T)$.

A coleção das bases já foi descrita acima e coincide com a coleção dos geradores de $\mathcal{M}(T)$ com cinco elementos. Este matroide tem 132 bases.

4.3 Forma recursiva para a construção das bases do matroide

Seja $T = (V, E)$ uma X -árvore. Dada uma aresta interior $\{u, v\} = f \in E$, a *contração* de T em f é a X -árvore denotada por T/f tal que:

- $V(T/f) = (V \setminus \{u, v\}) \cup \{\theta_f\}$, onde θ_f é um novo vértice que substitui u e v ;
- $E(T/f) = E \setminus \{f\}$, em que identificamos uma aresta $\{a, b\} \in E \setminus \{f\}$ com a aresta:
 - $\{a, b\} \in E(T/f)$ se $\{a, b\} \cap \{u, v\} = \emptyset$;
 - $\{a, \theta_f\} \in E(T/f)$ se $b \in \{u, v\}$;
 - $\{b, \theta_f\} \in E(T/f)$ se $a \in \{u, v\}$.

É fácil de verificar que T/f é ainda uma X -árvore e que se $g \in E$ é uma aresta interior de T diferente de f , então g é também uma aresta interior de T/f . Deste modo é possível formar as contrações $(T/f)/g$ e $(T/g)/f$, e verificar que $(T/f)/g = (T/g)/f$. Assim, dado $F \subseteq E$ um conjunto de arestas interiores de T , T/F designará a X -árvore que se obtém fazendo a contração sucessiva (por qualquer ordem) das arestas de F . Pela definição de contração que foi dada, $E(T/F)$ identifica-se naturalmente com $E \setminus F \subseteq E$. Por fim, observamos que dada uma X -árvore $T = (V, E)$, existe um conjunto F de arestas interiores de T tal que T/F é uma X -árvore estrela. Estas observações, conjugadas com a descrição do matroide $\mathcal{M}(T^*)$ dada na Proposição 1 para uma X -árvore estrela T^* , motivam o estudo da relação entre os matroides $\mathcal{M}(T)$ e $\mathcal{M}(T/f)$, que faremos de seguida.

Com o intuito de ilustrar a contração sucessiva das arestas interiores de uma X -árvore, vamos considerar a X -árvore $T = (V, E)$ representada na Figura 4.19.

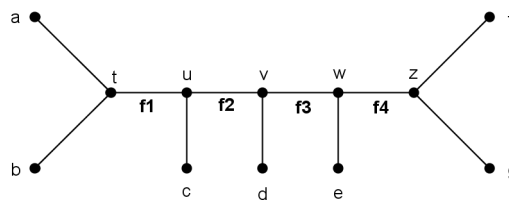


Figura 4.19: X -árvore T .

O conjunto das arestas interiores de $T = (V, E)$ é $\{f_1, f_2, f_3, f_4\}$ e o conjunto dos vértices interiores é $\{t, u, v, w, z\}$.

Começamos por contrair a aresta $f_1 = tu$ à X -árvore T . A X -árvore obtida, representada na Figura 4.20, é a X -árvore denotada por $T_1 = T \setminus \{f_1\}$, tal que:

- $V(T_1) = V(T \setminus \{f_1\}) = (V \setminus \{u, v\}) \cup \{\theta_{f_1}\}$,

- $E(T_1) = E(T \setminus \{f_1\}) = (E \setminus f_1)$,

onde θ_{f_1} é o novo vértice que substitui u e t .

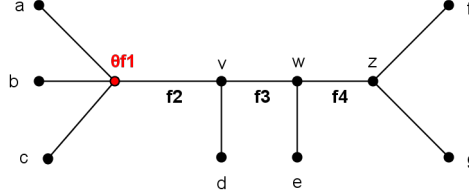


Figura 4.20: X -árvore $T_1 = T/\{f_1\}$.

Do mesmo modo, contraimos as arestas interiores f_2 , f_3 e f_4 , até obter a X -árvore estrela. Esta construção está representada na Figura 4.21.

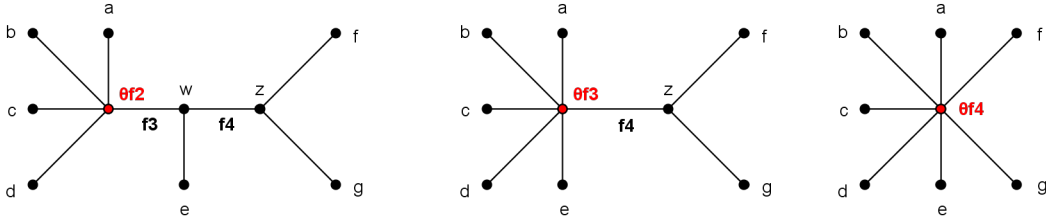


Figura 4.21: X -árvores: $T_2 = T_1/\{f_2\}$; $T_3 = T_2/\{f_3\}$; $T_4 = T_3/\{f_4\}$

Se F é um conjunto de arestas interiores de T , já observámos que $E(T/F) = E \setminus F \subseteq E$. Logo, a inclusão $\mathbb{R}^{E \setminus F} \hookrightarrow \mathbb{R}^E$ que se obtém estendendo $w : E \setminus F \rightarrow \mathbb{R}$ a E , definindo $w(f) = 0$ para todo o $f \in F$, dá origem a uma sobrejeção $\widehat{\mathbb{R}^E} \twoheadrightarrow \widehat{\mathbb{R}^{E \setminus F}}$, dada pela restrição

$$\lambda : \mathbb{R}^E \rightarrow \mathbb{R} \mapsto \lambda|_{E \setminus F} : E \setminus F \rightarrow \mathbb{R}.$$

Em particular, usaremos a identificação seguinte:

$$\mathbb{R}^{E \setminus F} = \{w \in \mathbb{R}^E : w(f) = 0 \text{ para todo o } f \in F\}.$$

Lema 5. Sejam $T = (V, E)$ uma X -árvore, $F \subseteq E$ um conjunto de arestas interiores de T e T/F a X -árvore obtida de T por contração das arestas em F . Então:

(i) Para todo o $xy \in \binom{X}{2}$, $\lambda_{xy}^T|_{E \setminus F} = \lambda_{xy}^{T/F}$.

(ii) Para todo o subconjunto $\mathcal{L} \subseteq \binom{X}{2}$, tem-se

$$\begin{aligned} \text{rank}^T(\mathcal{L}) &= \text{rank}^{T/F}(\mathcal{L}) + \dim\{\lambda \in \langle \mathcal{L} \rangle^T : \lambda(w_e) = 0, \forall e \in E \setminus F\} \\ &\leq \text{rank}^{T/F}(\mathcal{L}) + |F|. \end{aligned}$$

(iii) Se \mathcal{L} é um *edge-weight* lasso para T então \mathcal{L} é um *edge-weight* lasso para T/F .

Demonstração. Dados $x, y \in X$, o caminho em T/F entre x e y obtém-se do caminho em T entre x e y , eliminando neste último as arestas de $f \in F$ que nele ocorrem (e fazendo a respetiva substituição das extremidades de f pelo vértice θ_f). Logo, $\lambda_{xy}^T(w) = \lambda_{xy}^{T/F}(w)$ para todo o $w \in \mathbb{R}^{E \setminus F} = \{w \in \mathbb{R}^E : w(f) = 0, \forall f \in F\}$, o que prova (i).

Para mostrar (ii), tome-se $\mathcal{L} \subseteq \binom{X}{2}$ e considere-se o subconjunto $\Phi : \widehat{\mathbb{R}^E} \rightarrow \widehat{\mathbb{R}^{E \setminus F}}$ induzida pela inclusão $\mathbb{R}^{E \setminus F}$. Então,

$$\begin{aligned} \text{rank}^T(\mathcal{L}) &= \dim(\langle \mathcal{L} \rangle^T) = \dim(\langle \lambda_{xy}^T : xy \in \mathcal{L} \rangle) \\ &= \dim(\langle \Phi(\lambda_{xy}^T) : xy \in \mathcal{L} \rangle) + \dim(\text{Ker}\Phi \cap \langle \mathcal{L}^T \rangle). \end{aligned}$$

Ora, $\Phi(\lambda_{xy}^T) = \lambda_{xy|E \setminus F}^T = \lambda_{xy}^T$ e dado $\lambda \in \widehat{\mathbb{R}^E}$,

$$\begin{aligned} \lambda \in \text{Ker}\Phi &\Leftrightarrow \lambda_{E \setminus F} = 0 \\ &\Leftrightarrow \lambda(w) = 0 \forall w \in \mathbb{R}^{E \setminus F} \\ &\Leftrightarrow \lambda(w_e) = 0 \forall e \in E \setminus F \end{aligned}$$

já que $\{e_e : e \in E \setminus F\}$ é uma base de $\mathbb{R}^{E \setminus F}$.

Logo,

$$\begin{aligned} \dim(\langle \Phi(\lambda_{xy}^T) : xy \in \mathcal{L} \rangle) &= \dim(\langle \lambda_{xy}^{T \setminus F} : xy \in \mathcal{L} \rangle) \\ &= \text{rank}^{T \setminus F}(\mathcal{L}) \end{aligned}$$

e

$$\begin{aligned} \dim(\text{Ker}\Phi \cap \langle \mathcal{L} \rangle^T) &= \dim\{\lambda \in \langle \mathcal{L} \rangle^T : \lambda(w_e) = 0 \forall e \in E \setminus F\} \\ &\leq \dim(\text{Ker}\Phi) \\ &= \dim(\widehat{\mathbb{R}^E}) - \dim(\text{Im}\Phi) \\ &= |E| - \dim(\widehat{\mathbb{R}^{E \setminus F}}) \\ &= |E| - (|E| - |F|) = |F| \end{aligned}$$

A prova de (iii) é consequência imediata de (ii), já que \mathcal{L} é *edge-weight* lasso para T
 $\Leftrightarrow \text{rank}(\mathcal{L})^T = |E|$. \square

Proposição 2. Dados uma X -árvore $T = V, E$, uma aresta interior $f \in E$, $xy \in \binom{X}{2}$ e B uma base do matroide $\mathcal{M}(T/f)$, denotemos por $\rho_{xy} \in \mathbb{R}^B$ a única função que satisfaz $\lambda_{xy}^{T/f} = \sum_{b \in B} \rho_{xy}(b) \lambda_b^{T/f}$. Então $B \cup \{xy\}$ é uma base de $\mathcal{M}(T)$ se e só se $\sum_{b \in B} \rho_{xy}(b) \delta_{f|b} \neq \delta_{f|xy}$. Logo, o conjunto das bases de $\mathcal{M}(T)$ é dado por

$$\mathcal{B}(T) = \{B \cup \{xy\} : xy \in \binom{X}{2}, B \in \mathcal{B}(T/f) \text{ e } \sum_{b \in B} \rho_{xy}(b) \delta_{f|b} \neq \delta_{f|xy}\}.$$

Demonstração. Pelo lema anterior, sabemos que dada uma base B' de $\mathcal{M}(T)$, B' é gerador de $\mathcal{M}(T/f)$, logo existe $xy \in B'$ tal que $B = B' \setminus \{xy\}$ é base de $\mathcal{M}(T/f)$, o que mostra que as bases de $\mathcal{M}(T)$ são da forma $B \cup \{xy\}$, com B base de $\mathcal{M}(T/f)$ e $xy \in \binom{X}{2}$. Resta determinar para que escolhas de $B \in \mathcal{B}(T/f)$ e $xy \in \binom{X}{2}$ é que $B \cup \{xy\}$ é base de $\mathcal{M}(T)$.

Sejam então $xy \in \binom{X}{2}$ e B uma base de $\mathcal{M}(T/f)$. Em particular, $\text{rank}^{T/f}(B) = |B| = |E| - 1$. Pela parte (ii) do lema anterior temos que

$$\begin{aligned} B \cup \{xy\} \text{ é base de } \mathcal{M}(T) &\Leftrightarrow \text{rank}^T(B \cup \{xy\}) = |E| \\ &\Leftrightarrow \dim \Lambda_{B,xy} = 1 \\ &\Leftrightarrow \dim \Lambda_{B,xy} \neq 0 \\ &\Leftrightarrow \Lambda_{B,xy} \neq \{0\}, \end{aligned}$$

onde $\Lambda_{B,xy} = \{\lambda \in \langle B \cup \{xy\} \rangle^T : \lambda(w_e) = 0, \forall e \neq f\}$.

Seja $\lambda \in \langle B \cup \{xy\} \rangle^T$. Então existem $c \in \mathbb{R}$ e $\rho \in \mathbb{R}^B$ tais que $\lambda = c\lambda_{xy}^T + \sum_{b \in B} \rho(b)\lambda_b^T$, e $\lambda \in \Lambda_{B,xy}$ se e só se

$$0 = \lambda|_{E \setminus \{f\}} = c\lambda_{xy}^{T/f} + \sum_{b \in B} \rho(b)\lambda_b^{T/f}. \quad (4.1)$$

Seja então $\rho_{xy} \in \mathbb{R}^B$ dada por $\lambda_{xy}^{T/f} = \sum_{b \in B} \rho_{xy}(b)\lambda_b^{T/f}$. Como as funções $\lambda_b^{T/f}$, com $b \in B$, são linearmente independentes, resulta de (4.1) que $\rho(b) = -c\rho_{xy}(b)$, para todo o $b \in B$. Assim, $\Lambda_{B,xy} = \langle \lambda_{xy}^T - \sum_{b \in B} \rho_{xy}(b)\lambda_b^T \rangle$ e

$$\begin{aligned} \Lambda_{B,xy} \neq \{0\} &\Leftrightarrow \left(\lambda_{xy}^T - \sum_{b \in B} \rho_{xy}(b)\lambda_b^T \right) (w_f) \neq 0 \\ &\Leftrightarrow \sum_{b \in B} \rho_{xy}(b)\delta_{f|b} \neq \delta_{f|xy}, \end{aligned}$$

o que estabelece o resultado pretendido. \square

Para ilustrar a Proposição 2, vamos considerar a X -árvore representada na Figura 4.1 (ii), uma X -árvore T com conjunto de folhas $X := \{a, c, b, d\}$.

Neste caso, existe uma relação linear entre as funções λ_{xy}^T com $xy \in \binom{X}{2}$:

$$\begin{aligned} \lambda_{ac}^T + \lambda_{bd}^T &= \lambda_{ad}^T + \lambda_{bc}^T: \mathbb{R}^E \longrightarrow \mathbb{R} \\ w &\mapsto 2w(e_{ab|cd}) + \sum_{e \in E; e \neq e_{ab|cd}} w(e). \end{aligned}$$

Esta é a única relação linear em $\{\lambda_{xy}^T : xy \in \binom{X}{2}\}$. Assim, as bases do matroide $\mathcal{M}(T)$, associado a esta X -árvore, são os quatro conjuntos constituídos por subconjuntos de \mathcal{L} de $\binom{X}{2}$, com cinco elementos, que não contêm exactamente uma das quatro cordas ac, ad, bc, bd , ou equivalentemente, com $|\mathcal{L} \cap \{ac, ad, bc, bd\}| \leq 3$. Há quatro folhas, logo $\binom{4}{2} = 6$ cordas, e destas seis cordas queremos construir conjuntos de cinco elementos, $\binom{6}{5} = 6$ hipóteses. Nestas seis hipóteses temos que retirar os conjuntos em que as quatro cordas ac, bd, ad, bc pertencem em simultâneo (existem dois conjuntos deste tipo). Na Figura 4.22, estão representados os quatro grafos $\Gamma(\mathcal{L})$, associados aos lassos \mathcal{L} de $\binom{X}{2}$, que são bases do matroide $\mathcal{M}(T)$.

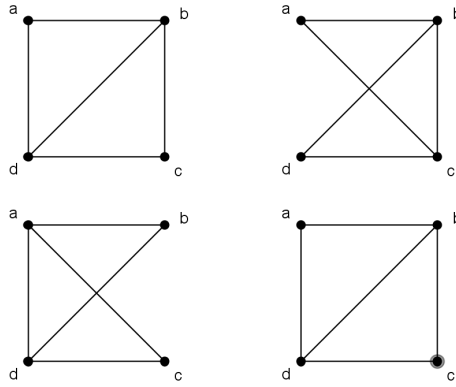


Figura 4.22: Grafos $\Gamma(\mathcal{L})$, associados às bases do matroide $\mathcal{M}(T)$.

Pelo Teorema 2.5, sabemos que, se \mathcal{L} é um *edge-weight* lasso, então o grafo $\Gamma(\mathcal{L})$ deve ser fortemente não bipartido com um ciclo de comprimento ímpar.

Claramente, se f coincide com a única aresta interior de T , isto é, a aresta denotada por $e_{ab|cd}$ da Figura 4.1 (ii), $T \setminus f$ é equivalente à árvore estrela T^* representada na Figura 2.6 (i). O grafo $\Gamma(\mathcal{L})$, correspondente às bases da X -árvore T^* , é o mínimo grafo fortemente não bipartido com conjunto de vértices X , que tem que ter um ciclo de comprimento ímpar. O grafo tem quatro cordas, logo o único ciclo de comprimento ímpar é um ciclo de comprimento três, neste caso, existem quatro possibilidades, ilustradas na Figura 4.23.

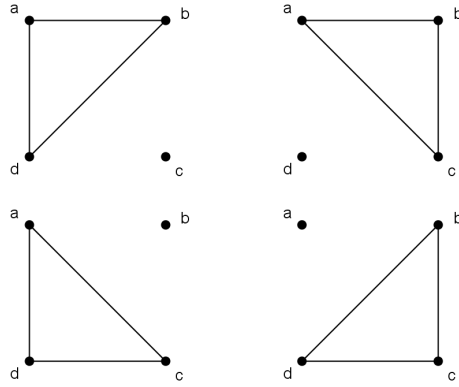


Figura 4.23: Grafos $\Gamma(\mathcal{L})$, associados às bases do matroide $\mathcal{M}(T)$.

O vértice que fica de fora tem três possibilidades de se ligar ao triângulo criado, ou seja, temos doze bases para o matroide associado à X -árvore estrela T^* . Os grafos $\Gamma(\mathcal{L})$ associados aos subconjuntos \mathcal{L} , que são bases do matroide associado à X -árvore estrela T^* , estão representados na Figura 4.12.

Observando as simetrias da árvore T , podemos formar duas órbitas relativas ao grupo de simetria de T , que são as bases:

$$B_1 := \{ab, bc, ca, da\} \text{ e } B_2 := \{ab, bc, ca, dc\},$$

representadas na Figura abaixo.

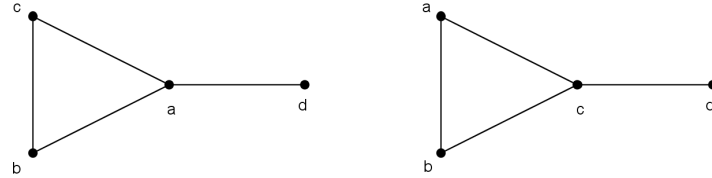


Figura 4.24: $\Gamma(B_1); \Gamma(B_2)$.

Com o intuito de encontrar as bases do matroide associado à X -árvore T , temos que encontrar os conjuntos $B = B_i \cup \{xy\}$ tais que B é uma base de T^* .

No caso de B_1 , as únicas cordas que ficaram de fora são db e dc . Vamos verificar se é possível acrescentar à base B_1 alguma das cordas.

Temos que $f := e_{ab|cd}$ e que:

$$\lambda_{db}^{T \setminus f} = \lambda_{da}^{T \setminus f} - \lambda_{ac}^{T \setminus f} + \lambda_{cb}^{T \setminus f} \text{ e } \lambda_{dc}^{T \setminus f} = \lambda_{da}^{T \setminus f} - \lambda_{ab}^{T \setminus f} + \lambda_{bc}^{T \setminus f}$$

enquanto,

$$\begin{aligned} \lambda_{db}^T(w_f) &= \lambda_{da}^T(w_f) - \lambda_{ac}^T(w_f) + \lambda_{cb}^T(w_f) \Leftrightarrow \delta_{f|bd} = \delta_{f|da} - \delta_{f|ac} + \delta_{f|cb} \Leftrightarrow \\ &\Leftrightarrow 1 - 1 + 1 = 1 \Leftrightarrow 1 = 1 \end{aligned}$$

e

$$\lambda_{dc}^T(w_f) = \lambda_{da}^T(w_f) - \lambda_{ab}^T(w_f) + \lambda_{bc}^T(w_f) \Leftrightarrow \delta_{f|dc} = \delta_{f|da} - \delta_{f|ab} + \delta_{f|bc} \Leftrightarrow \\ \Leftrightarrow 0 = 1 - 0 + 1 \Leftrightarrow 0 \neq 2$$

o que implica, que para todas as bases B_1 , podemos adicionar cordas dc , mas não db .

No caso de B_2 , as únicas cordas que ficaram de fora são da e db . Vamos verificar se é possível acrescentar à base B_2 alguma destas cordas:

$$\lambda_{da}^{T \setminus f} = \lambda_{dc}^{T \setminus f} - \lambda_{cb}^{T \setminus f} + \lambda_{ba}^{T \setminus f} \text{ e } \lambda_{db}^{T \setminus f} = \lambda_{dc}^{T \setminus f} - \lambda_{ca}^{T \setminus f} + \lambda_{ab}^{T \setminus f}$$

enquanto,

$$1 = \lambda_{da}^T(w_f) \neq \lambda_{dc}^T(w_f) - \lambda_{cb}^T(w_f) + \lambda_{ba}^T(w_f) = -1$$

e

$$1 = \lambda_{db}^T(w_f) \neq \lambda_{dc}^T(w_f) - \lambda_{ca}^T(w_f) + \lambda_{ab}^T(w_f) = -1$$

o que implica que, para as bases do tipo B_2 , podemos adicionar qualquer uma das cordas.

Exemplo 22. Com o intuito de descobrir uma base do matroide associado à X -árvore representada na Figura 4.25, vamos usar a construção recursiva, enunciada na Proposição 2.

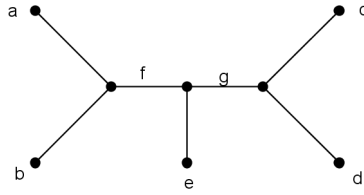


Figura 4.25: X -árvore com cinco folhas.

1) Uma base do matroide associado à X -árvore $T/\{f, g\}$, ou seja, a árvore estrela representada na Figura 4.26.

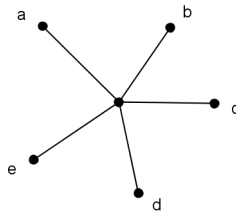


Figura 4.26: $T/\{f, g\}$, árvore estrela com cinco folhas: T^*

Como já vimos anteriormente, as bases do matroide $\mathcal{M}(T/\{f, g\}) = \mathcal{L}(T^*)$, são conjuntos \mathcal{L} para os quais o seu grafo associado, $\Gamma(\mathcal{L})$ tem uma das seguintes representações:

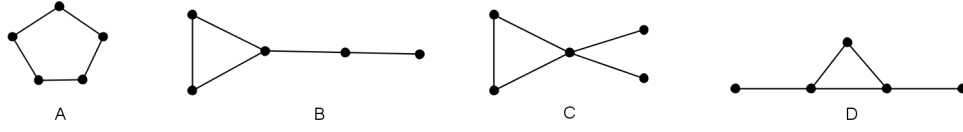


Figura 4.27: Grafos $\Gamma(\mathcal{L})$, associados a um lasso \mathcal{L} que são bases do matroide associado à X -árvore da Figura 4.26.

Por exemplo, o lasso $\mathcal{L} := \{ab, bc, bd, cd, de\}$, que é uma configuração D , é uma base do matroide $\mathcal{M}(T^*)$.

2) Uma base do matroide associado à X -árvore $T/\{g\}$, ou seja, à árvore representada na Figura 4.28.

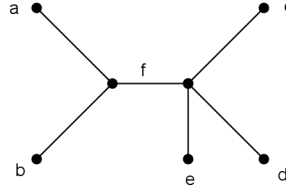


Figura 4.28: $T/\{g\}$.

Seja xy uma corda e f a aresta interior da X -árvore. Pela Proposição 2 sabemos que, se B é uma base de $T/\{f, g\}$ e xy uma corda, então $B \cup \{xy\}$ é uma base do matroide associado a $T/\{g\}$ se e só se:

$$\sum_{b \in B} \rho_{xy}(b) \delta_{f|b} \neq \delta_{f|xy}.$$

Neste caso, as cordas xy , podem ser: ac, ad, ae, be ou ce .

Para verificar se $B \cup \{xy\}$ é uma base de $T/\{g\}$, basta verificar se a desigualdade se verifica.

Seja $B = \{ab, bc, bd, cd, de\}$ uma base de T^* .

$$\lambda_{ac}^{T/\{f,g\}} = \lambda_{ab}^{T/\{f,g\}} - \lambda_{bd}^{T/\{f,g\}} + \lambda_{cd}^{T/\{f,g\}}$$

então

$$\begin{aligned} \lambda_{ac}^{T/\{g\}} &= \lambda_{ab}^{T/\{g\}} - \lambda_{bd}^{T/\{g\}} + \lambda_{cd}^{T/\{g\}} \Leftrightarrow \\ \Leftrightarrow \delta_{f|ac} &= \delta_{f|ab} - \delta_{f|bd} + \delta_{f|cd} \Leftrightarrow \\ \Leftrightarrow 1 &= 0 - 1 + 0 \Leftrightarrow \\ \Leftrightarrow 1 &\neq 1 \end{aligned}$$

Logo $B \cup \{ac\} = B_1 = \{ab, bc, bd, cd, de, ac\}$ é uma base de $T/\{g\}$.

3) Uma base do matroide associado à X -árvore T , a árvore representada na Figura 4.25.

Pela Proposição 2 sabemos que, se B_1 é uma base de $T/\{g\}$ e xy uma corda, então $B_1 \cup \{xy\}$ é uma base do matroide associado a T se e só se:

$$\sum_{b \in B} \rho_{xy}(b) \delta_{f|b} \neq \delta_{f|xy}.$$

Neste caso, as cordas xy , podem ser: ad, ae, be ou ce onde $B_1 = \{ab, bc, bd, cd, de, ac\}$.

$$\lambda_{ad}^{T/\{g\}} = \lambda_{ae}^{T/\{g\}} - \lambda_{bc}^{T/\{g\}} + \lambda_{bd}^{T/\{g\}} + \lambda_{cd}^{T/\{g\}} - \lambda_{de}^{T/\{g\}}$$

então

$$\begin{aligned} \lambda_{ad}^T &= \lambda_{ae}^T - \lambda_{bc}^T + \lambda_{bd}^T + \lambda_{cd}^T - \lambda_{de}^T \Leftrightarrow \\ \Leftrightarrow \delta_{f|ad} &= \delta_{f|ae} - \delta_{f|bc} + \delta_{f|bd} + \delta_{f|cd} - \delta_{f|de} \Leftrightarrow \\ \Leftrightarrow 1 &= 0 - 1 + 1 + 0 - 1 \Leftrightarrow \\ &\Leftrightarrow 1 \neq -1 \end{aligned}$$

Logo, $B_1 \cup \{ad\} = B_2 = \{ab, bc, bd, cd, de, ac, ad\}$ é uma base de $\mathcal{M}(T)$.

A árvore T tem os seis seguintes grupos de simetria: $\{Id, (ab), (cd), (ab)(cd), (ac)(bd), (ad)(bc)\}$. Se aplicarmos as simetrias à base B_2 , obtemos cinco bases distintas:

$$B_3 = \{ab, ac, ad, cd, de, bc, bd\}$$

$$B_4 = \{ab, bd, bc, cd, ce, ad, ac\}$$

$$B_5 = \{ab, ad, ac, cd, ce, bd, bc\}$$

$$B_6 = \{cd, da, bd, ab, be, ac, cb\}$$

$$B_7 = \{dc, bc, ca, ba, ae, db, ad\}$$

que são bases do matroide associado à X -árvore T .

4.4 Caraterização da X -árvore pelo seu matroide

Vejamos que o matroide $\mathcal{M}(T)$ de uma X -árvore T determina essa X -árvore, a menos de isomorfismos de X -árvores.

Definição 25. Dada uma X -árvore T e $Y \subseteq X$ com $|Y| \geq 3$, a *restrição de T a Y* é a Y -árvore denotada por $T|_Y$ que se obtém da sub-árvore de T minimal que contém os caminhos entre os vértices de Y , por supressão dos vértices de grau 2 (isto é, substituindo cada vértice v de grau 2 e as duas arestas incidentes com v por uma única aresta com extremidades nos dois vértices adjacentes a v).

O conjunto de vértices e arestas de $T|_Y$ denotam-se por V_Y e $E|_Y$.

Dado w uma função peso nas arestas de T , $w|_Y$ é a função peso induzida em $T|_Y$ que atribui a cada aresta $\{u, v\}$ de $E|_Y$ a distância em T entre os vértices u e v , relativamente a w .

Note-se que, para $y, y' \in Y$ e $w \in \mathbb{R}^E$, $\lambda_{yy'}^T(w) = \lambda_{yy'}^{T|_Y}(w|_Y)$

Lema 6. Dada uma X -árvore T e $Y \subseteq X$ um conjunto com quatro elementos, digamos $Y = \{a, b, c, d\}$, $T|_Y \cong T_{ab|cd} \Leftrightarrow$ existe alguma base B de $\mathcal{M}(T)$ que contém $\mathcal{L}_{abcd} = \{ab, bc, cd, ad\}$ e nenhuma base de $\mathcal{M}(T)$ contém $\{ac, ad, bc, bd\}$.

Demonstração. Se $T|_Y \cong T_{ab|cd}$, então as funções $\lambda_{xy}^{T|_Y}$ com $xy \in \mathcal{L}_{abcd}$ são linearmente independentes, logo o mesmo se sucede com a função λ_{xy}^T , $xy \in \mathcal{L}_{abcd}$.

Assim, existe alguma base de $\mathcal{M}(T)$ que contém \mathcal{L}_{abcd} . Além disso, em $T_{ab|cd}$,

$$\lambda_{ac}^{T|_Y} + \lambda_{bd}^{T|_Y} = \lambda_{ad}^{T|_Y} + \lambda_{bc}^{T|_Y},$$

logo em T ,

$$\lambda_{ac}^T + \lambda_{bd}^T = \lambda_{ad}^T + \lambda_{bc}^T,$$

e portanto nenhuma base de $\mathcal{M}(T)$ contém $\{ac, ad, bc, bd\}$.

Reciprocamente, das quatro possibilidades para $T|_Y$: $T_{ab|cd}$, $T_{ad|bc}$, $T_{ac|bd}$ e T^* , a árvore estrela com folhas $\{a, b, c, d\}$. a relação

$$\lambda_{ab}^{T'} + \lambda_{cd}^{T'} = \lambda_{ad}^{T'} + \lambda_{bc}^{T'}$$

é satisfeita em $T' = T_{ac|bd}$ e em $T' = T^*$, logo, se \mathcal{L}_{abcd} está contido em alguma base de $\mathcal{M}(T)$, resulta que $T|_Y \not\cong T_{ab|cd}$ e $T|_Y \not\cong T^*$. Além disso, se $T|_Y \cong T_{ad|bc}$ então $\{\lambda_{ac}^{T|_Y}, \lambda_{ad}^{T|_Y}, \lambda_{bc}^{T|_Y}, \lambda_{bd}^{T|_Y}\}$ é linearmente independente, o que implica que $\lambda_{ac}^T, \lambda_{ad}^T, \lambda_{bc}^T, \lambda_{bd}^T$ também são linearmente independentes, portanto existe alguma base de $\mathcal{M}(T)$ que contém $\{ac, ad, bc, bd\}$, o que é absurdo. Logo, $T|_Y \not\cong T_{ad|bc}$.

Logo $T|_Y \cong T_{ab|cd}$. □

Definição 26. Dado uma X -árvore T , definimos $Q(T) = \{ab|cd : \{a, b, c, d\} \in \binom{X}{4} \text{ e } T|_{\{a,b,c,d\}} \cong T_{ab|cd}\}$.

Teorema 4.3. ([3, Corolário 6.3.8]) Dadas duas X -árvores T_1 e T_2 , $T_1 \cong T_2 \Leftrightarrow Q(T_1) = Q(T_2)$.

Corolário 2. Para quaisquer X -árvores T_1 e T_2 , $\mathcal{M}(T_1) = \mathcal{M}(T_2) \Leftrightarrow T_1 \cong T_2$.

Demonstração. (\Leftarrow .) Se $T_1 \cong T_2$ então é fácil ver que $\mathcal{M}(T_1) = \mathcal{M}(T_2)$, já que um isomorfismo de X -árvores se restringe à identidade em X .

(\Rightarrow .) Como $\mathcal{M}(T_1) = \mathcal{M}(T_2)$ então as bases dos dois matroides são iguais. Pelo Lema 6, sabemos que se $\mathcal{M}(T_1) = \mathcal{M}(T_2)$ então $Q(T_1) = Q(T_2)$ e pelo Teorema 4.3 podemos concluir que $T_1 \cong T_2$. □

Capítulo 5

Conclusão

O estudo das matérias contidas neste trabalho, salienta o papel relevante da Matemática e dos seu métodos na reconstrução filogenética.

A filogenética, apresenta uma limitação que impossibilita a construção da árvore evolutiva de certos conjuntos de indivíduos, por falta de informação relativa aos mesmos. Os artigos estudados nesta dissertação, lidam com esta limitação e apresentam uma solução para este problema. Foi possível, com certos conjuntos de pares de folhas (indivíduos), determinar a árvore evolutiva de um conjunto de indivíduos.

Depois de estudada a teoria dos lassos, nomeadamente a forma como reconstruir uma árvore filogenética à custa destes, foi estudada a teoria de matroides. Esta última teoria permitiu a associação de uma estrutura combinatória, um matroide, às árvores filogenéticas.

Considero atingidos com êxito todos os objetivos aos quais me propus aquando do início da realização desta dissertação.

Sendo o principal objetivo, a caracterização do matroide $\mathcal{M}(T)$ associado a uma árvore filogenética.

Bibliografia

- [1] Oxley, J., *Matroid Theory*, Second edition, Oxford University Press, USA, 2001.
- [2] Welsh, D. A., *Matroid theory*, Academic Press, New York, 1976.
- [3] Semple, C. and Steel, M. *Phylogenetics*, Oxford University Press, 2003.
- [4] Dress, A., Huber, K. and Steel, M., "*Lassoing*" a phylogenetic tree I: *Basic properties, shellings and covers*, J. Math. Biol, 2011. ISSN: 0303-6812 (Print) 1432-1416 (Online), retrieved from <http://arxiv.org/pdf/1102.0309.pdf>.
- [5] Dress, A., Huber, K. and Steel, M., *A matroid associated with a phylogenetic tree*, 2013. Discrete Mathematics and Theoretical Computer Science, Vol 16, No 2, retrieved from <http://arxiv.org/pdf/1307.7287.pdf>.
- [6] White, N., *Theory of Matroids*, University of Florida, 1986.
- [7] Gregório, L., Ribeiro S. C., *Álgebra Linear*, Mc Graw-Hill, 1985.
- [8] Wilson, W. R., *Introduction to Graph Theory*, third edition, Longman Scientific and Technical, 1985.
- [9] Rocha de la Féria, R. C. S., *Reconstrução filogenética*, Faculdade de Ciências da Universidade do Porto, 2008.
- [10] Balakrishnan, J., Ranganathan, K. *A Textbook of Graph Theory*. Springer-Verlag (1999).